Systems Biology Approaches To Unravel The Mechanisms Underlying The Evolution Of The Transcription Factor Brachyury

Thesis

SYSTEMS BIOLOGY APPROACHES TO UNRAVEL THE MECHANISMS UNDERLYING THE EVOLUTION OF THE TRANSCRIPTION FACTOR BRACHYURY

Jovana Ranđelović

Doctor of Philosophy

School of Life, Health and Chemical Sciences

Affiliated Research Centre (ARC):

Stazione Zoologica Anton Dohrn

Naples, Italy

The Open University

November 2020

This thesis work has been carried out in the laboratory of Dr. Giovanna Benvenuto and Dr. Maria I. Arnone at the Stazione Zoologica Anton Dohrn, Naples, Italy

Director of Studies: Dr. Giovanna Benvenuto, Stazione Zoologica Anton Dohrn, Naples, Italy

Internal supervisor: Dr. Maria I. Arnone, Stazione Zoologica Anton Dohrn, Naples, Italy

External Supervisor: Prof. José Luis Gómez-Skarmeta, Centro Andaluz de Biología del Desarrollo, CSIC/Universidad Pablo de Olavide, Seville, Spain

ABSTRACT

Bilaterians share a conserved molecular toolkit depicted in Gene Regulatory Networks (GRNs) that controls embryonic development and morphogenesis. Modifications of this toolkit can lead to the evolution of new cell types and new body plans. This work investigates the role of transcription factor (TF) Brachyury in the embryonic development of two sea urchin species – *Strongylocentrotus purpuratus* and *Paracentrotus lividus*.

In non-vertebrate and vertebrate chordates, Brachyury acts as an activator of mesodermal genes. However, in protostomes, Brachyury activates endodermal and ectodermal genes, suggesting its ancestral function. Sea urchin, due to its phylogenetic position as a non-vertebrate deuterostome, the thorough characterization of its GRNs, and the expression of Brachyury in ectodermal and endodermal domains, served as a model system for this study. The goal of this study is to identify direct and indirect Brachyury targets and untangle the similarities and differences in Brachyury's role and function in the early development of two sea urchin species by combining gene perturbation and high-throughput sequencing technologies. The knock-down of Brachyury, followed by differential transcriptomics, allowed the investigation of its putative targets. Moreover, the use of available Brachyury ChIP-seq data to identify Brachyury direct interactions and ATAC-seq to discover open chromatin regions has led to the extension of a known GRN around Brachyury in *S. purpuratus*.

This work showed that Brachyury acts mostly as an ectodermal and endodermal activator in both sea urchin species. Some mesodermal genes detected after the perturbation were commonly affected in both species, although it appears that Brachyury can directly repress mesodermal fate in endodermal veg2 cell lineage in *S. purpuratus*. The self-regulatory feedback loop of Brachyury seems to be conserved. Moreover, the analysis of the obtained data suggested the direction for further work.

To my family

ACKNOWLEDGMENTS

Without all of you, this would not have been possible.

Ina and Giovanna, thank you for hosting me in your lab and being my guidance and excellent mentors during this three-year-long journey.

Special thanks go to Danila and Periklis for being able to keep up with me every day before, during, and after the time we spent in the lab. Also, the weekends. Thank you for your enormous help and dedication to my project, and thank you for becoming one of my closest friends. I will never forget our aperitifs, pizza nights, Russian dinners, aimless Sunday walks around Naples, you suffering from my contemporary art museum choices, long train rides to Salerno coast, countless cigarette and coffee breaks, melanzane e patate (not a euphemism), gossip sessions, Danila's strange cocktail choices, Russian proverbs…should I stop now?

Thank you, Solenn, for always pushing me to continue when it was hard and for believing in me. You are truly one of the kindest persons I've met! Thank you, Alfonsina and Miriam, my true Neapolitan family. Without you, I would have been 10 kg lighter and not need to buy new clothes every summer. Our trips to the beach, wine, and sushi nights, Barcelona…they won't be forgotten!

Thank you, Natalia, you made my five-month stay in Naples turn into another three years.

Thank you, Mima, Danica, Margareta, Irena, and Maja, for being in touch with me almost every day for all those long five years I had been far from home. Thanks, Nikola, for hosting me and feeding me any time I needed a break from all the flights and long layovers before reaching my two homes.

Thanks to all the wonderful colleagues and friends I met in the lab: Ines, Marie-Lyne, Maria, Elijah, Mena, Ylenia, Ivan, Monika, and many, many more…you were all my support!

Thanks to all the people from Bioinforma that taught me how to work on the cluster, and of course, for saving my laptop after it burnt. Thank you, Davide, for taking care of the sea urchins.

Thank you, Jose-Luis, for insightful conversions and pieces of advice. Thanks, Marta, for a great time during our experiments and time outside of the lab.

I would love to thank my examiners, Dr. Uli Technau and Dr. Gabriele Procaccini, for participating in a great and interesting thesis discussion. Thank you for all your suggestions for my thesis improvement.

Thanks to all my friends around the world who inspired me to grow and who stayed in touch with me all this time.

Finally, the biggest thanks go to my family: Hvala mama, hvala tata, hvala Jasna! Thank you for believing in me and never stopping me from achieving my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter focuses on the general introduction to the thesis. It describes the evolution of metazoan germ layers, an overview of the evolutionary changes of the expression patterns of the transcription factor Brachyury, gene regulatory networks, the experimental models used in this study, and the aims of this thesis.

1.1 Metazoan germ layers evolution

Different germ layers that emerge during early animal development and from which all organs and tissues develop were first identified and described during the 19th century by Karl Ernst von Baer (Baer, 1828; Oppenheimer, 1990). The layers were given the names ectoderm, endoderm, and mesoderm, based on their localization during gastrulation and adult structures that develop from them. For instance, in bilaterians, ectoderm giving rise to the protective outer layer (skin) and nervous system, endoderm giving rise to the digestive system, and mesoderm giving rise to the heart, muscles, and skeleton. Gastrulation can be defined as the process during which a subset of cells migrates to the interior of the single germ layer structure (blastula) to form primary endoderm. Mesoderm formation is sometimes in timely concordance with the process of endoderm formation, although some animals develop mesoderm earlier or later, or they lack it entirely (diploblastic animals like sponges and cnidarians). The discovery and description of ectoderm, endoderm, and mesoderm marked the beginning of the “germ-layer theory” (Baxter, 1977). Later, Thomas Henry Huxley developed the theory further by introducing the homology concept. Huxley believed that the two germ layers of coelenterates (Cnidaria and Ctenophora) shared common physiological characteristics with the ectoderm and endoderm of vertebrate embryos (Baxter, 1977; Gilbert, 2003). Charles Darwin’s theory of evolution marked a milestone that transformed

comparative embryology into evolutionary embryology. Gilbert has described four men as the founding fathers of evolutionary embryology: Fritz Müller, who connected natural selection and embryology when proclaiming that Nauplius larva is the common source of all crustaceans; Alexander Kowalevsky, who was the first to describe the gastrulation process and revived Huxley's homology theory; Ernst Haeckel, who proclaimed the famous "Ontogeny recapitulates phylogeny," or that the origin of new species followed the same rules as the origin of new embryonic structures; and Francis M. Balfour, who inferred that groups that share common larvae share common ancestors (Gilbert, 2003).

What started as comparative embryology diverged into many new fields such as evolutionary developmental biology (Evo-Devo), ecological evolutionary developmental biology (Eco-Evo-Devo), and medical developmental biology. Through a period longer than 200 years, classic embryology incorporated everything that came along the way: evolutionary biology, genetics, ecology, paleontology, medicine, molecular biology, bioinformatics, statistics, and computational modeling. This thesis work is still within the boundaries of evolutionary developmental biology, or Evo-Devo. Even though this field has changed immensely since it originally started in the nineteenth century, the primary goal stayed the same: to decipher the relationships between the three different germ layers and their evolution.

All multicellular organisms belong to the monophyletic group known as Metazoa, which descended from the last common ancestor, also named Urmetazoa (Müller, 2001; Sebé-Pedrós, Degnan and Ruiz-Trillo, 2017). Metazoans can be subdivided into non-bilaterian (sponges, placozoans, ctenophores, and cnidarians) and bilaterian (protostomes and deuterostomes). Non-bilaterian groups are characterized by the lack of bilateral body symmetry and the mesoderm germ layer. Porifera (sponges), Placozoa, Ctenophora (comb jellies), and Cnidaria (jellyfish, anemones, and corals) are grouped artificially and are also known as early-branching or basal metazoans (Dohrmann and Worheide, 2013). Sponges lack typical body symmetry as adults, but they have radial symmetry during their larval stages (Genikhovich and Technau, 2017). Placozoans consist of a single species named *Trichoplax adhaerens* characterized by their simple body plan lacking

anterior-posterior polarity, gut, nerve, or muscle cells. In fact, it has been shown that Placozoa only possesses six cell types (Smith *et al.*, 2014). Their embryonic development is almost unknown (Eitel *et al.*, 2011). Ctenophores (comb jellies) are non-bilaterian metazoans of unclear phylogenetic position. Their body plan has bi-radial symmetry and is a complex collection of autapomorphies, such as the comb rows of cilia. Comb jelly embryology is unique, yet it exhibits bilaterian-like traits (Jager and Manuel, 2016; Genikhovich and Technau, 2017). Comb jellies possess structures such as the apical organ, mesodermal musculature, and the tube gut with both mouth and anus, all features that put them outside of Porifera, Placozoa, and Cnidaria, even though they are non-bilaterian animals. It is believed that these are the results of convergent evolution (Jager and Manuel, 2016; King and Rokas, 2017). Cnidarians have both radial (jellyfish and hydra) and bilateral symmetry (corals, sea anemones, and sea pens, known as Anthozoa). Anthozoans are of particular interest as a possible link to reconstructing the last common bilaterian ancestor (Genikhovich and Technau, 2017).

The mesoderm evolution was the defining timepoint in the emergence of triploblastic animals or Bilateria about 600 million years ago (Chen *et al.*, 2004). Bilaterian groups are characterized by bilateral symmetry, or two orthogonal body axes and three germ layers. The classical interpretation of the mouth and anus formation during gastrulation led to the recognition of two major groups of Bilateria, the Protostomia and Deuterostomia. In the case of Protostomia (from Greek πρώτο - *first*, and στόμα - *mouth*), the mouth is formed from the blastopore during embryonic development, while in Deuterostomia (from Greek δεύτερο - *second*, and στόμα - *mouth*), the anus is formed from the blastopore and the mouth forms later. Currently, Protostomia is divided into Spiralia (mollusks, annelids, platyhelminths), characterized by the spiral cleavage during development, and Ecdysozoa (nematodes, priapulids, arthropods), characterized by ecdysis or molting (Genikhovich and Technau, 2017). Ambulacraria (Echinodermata and Hemichordata) is a sister group to Chordata (Urochordata, Cephalochordata, and Vertebrata), and together they make up Deuterostomia (Genikhovich and Technau, 2017). However, this boundary is not entirely clear because some animals that are traditionally classified as protostomes show a deuterostome developmental scenario, e.g., ecdysozoan brachiopod

species *Novocrania anomala* (Martín-Durán *et al.*, 2016). Furthermore, there is strong evidence that blastopore develops into mouth and anus in Protostomia, and into anus only in Deuterostomia (Arendt, Technau and Wittbrodt, 2001).

Blastopore can be described as the organizing center of gastrulation. In both diploblastic and triploblastic animals, it marks the region from which the ectodermal cells start to move interiorly and form the primitive gut. As mentioned already, gastrulation is connected to the mesoderm formation as well. Moreover, blastopore is the place that marks the regional separation of the ectoderm and endoderm (Technau and Scholz, 2003). Mesoderm emergence played an essential step in the radiation of Bilateria and gave a chance for complex body plans and organs to evolve (Pérez-Pomares and Muñoz-Chápuli, 2002). Since mesoderm was the evolutionary novelty, it is of great importance to detangle which genes are involved in mesoderm specification (Martindale, Pang and Finnerty, 2004). First, one should find the relationship between the known gene regulation of the mesoderm formation in bilaterians and look for the same toolkit in non-bilaterians. If the new structures emerge by "tuning" the old structures, one should look into the "pre-mesoderm" animals. Today we know, mostly from the gene regulation studies in vertebrates, which transcription factor families and signaling molecules are essential parts of the mesoderm formation toolkit: T-box (*brachyury*), MADS-box (*mef2*), bHLH (*twist*), and zinc finger (GATA factors and *snail*). These transcription factor families play a crucial role in bilaterian mesoderm and endomesoderm formation (Spring *et al.*, 2002; Technau and Scholz, 2003). These TFs do appear in non-bilaterian genomes, as shown by various groups, who studied "mesodermal" gene expression in the diploblastic animals: the sea anemone *Nematostella vectensis*, the jellyfish *Podocoryne carnea*, and the freshwater polyp *Hydra* (Technau, 2001; Spring *et al.*, 2002; Scholz and Technau, 2003). The mentioned studies prove that the toolkit for mesoderm differentiation existed before the mesoderm invention, albeit used differently. Evolutionary novelties can arise from duplicated genes in three different ways:

- sub-functionalization (when gaining a complementary function)
- neo-functionalization (when gaining a new function)

- non-functionalization (when one gene copy gets lost, and the remaining copy retracts its primary function) (Force *et al.*, 1999).

One of the most interesting groups of transcription factors to look at is the T-box transcription factor family because of their involvement in animal development regulation. *Brachyury* is the founding member of this group, and it has been described as a crucial gene in mesoderm formation in vertebrates after its discovery (Herrmann *et al.*, 1990; Wilkinson, Bhatt and Herrmann, 1990). The next subchapter gives a closer look at Brachyury's structure, function, and evolution.

1.2 Brachyury and T-box transcription factor family

Brachyury was first discovered in developing mouse embryos by Herrmann and collaborators (Herrmann *et al.*, 1990; Wilkinson, Bhatt and Herrmann, 1990). T-box transcription factor family is a key family involved in the development of metazoans (Smith, 1999; Papaioannou, 2001, 2014; Wilson and Conlon, 2002; Showell, Binder and Conlon, 2004). However, T-box genes originated in pre-metazoans. These genes are present in the genomes of early-branching fungi, ichthyosporeans, and unicellular holozoans (Mendoza, Taylor and Ajello, 2002; Sebé-Pedrós, Ariza-Cosano, *et al.*, 2013; Suga *et al.*, 2013; Sebé-Pedrós and Ruiz-Trillo, 2017). T-box transcription factor family is characterized by a conserved T-box DNA binding domain of 180–200 amino acids. Even though the sequence variation within the T-box family members exists, all family members bind to the TCACACCT DNA consensus sequence (Wilson and Conlon, 2002). Crystallographic studies have shown the structure of the Brachyury protein (Figure 1.1)(Müller *et al.*, 1997). The DNA-binding domain of Brachyury is represented by the first 229 amino acids of the protein that binds to the 20 nucleotides-long partially palindromic sequence T[G/C]ACACCTAGGTGTGAAATT (Kispert and Herrmann, 1993) when the TF is in a dimeric form (Müller *et al.*, 1997).

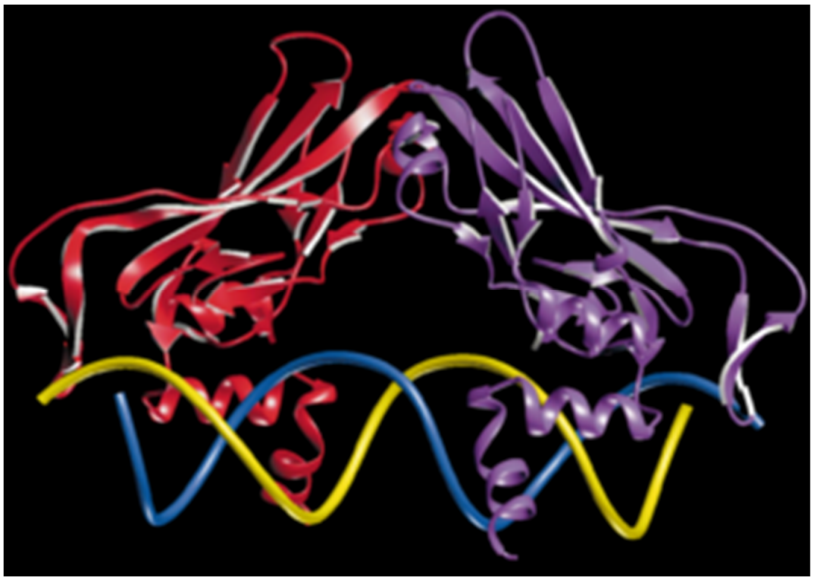


Figure 1.1 Three-dimensional diagram of the *Xenopus laevis* Brachyury (XBra) T-domain (residues 39-221) bound to a palindromic binding site, shown in dimeric form. The two DNA strands are shown in the blue and yellow. (Adapted and modified from Smith, 1999)

The DNA-binding domain location is in the N-terminal portion of the protein, and transcriptional activation domain or activation and repression domains are in the C-terminal region (Papaioannou, 2014). Wilson and Conlon's study showed some evidence that only a short sequence belonging to the T-box domain determines the specificity of a given T-box transcription factor (Wilson and Conlon, 2002; Papaioannou, 2014). For instance, in *Xenopus laevis*, the specificity of T-box TF members is determined by just one amino acid residue at position 149. XBra (*Xenopus* Brachyury orthologue) differs from Vegetally localized transcription factor (VegT) and Eomesodermin (Eomes) by a single lysine (Figure 1.2A). Moreover, comparing Brachyury protein sequences from different species, they all show the presence of the lysine residue, as shown in Figure 1.2B (Wilson and Conlon, 2002; Papaioannou, 2014).

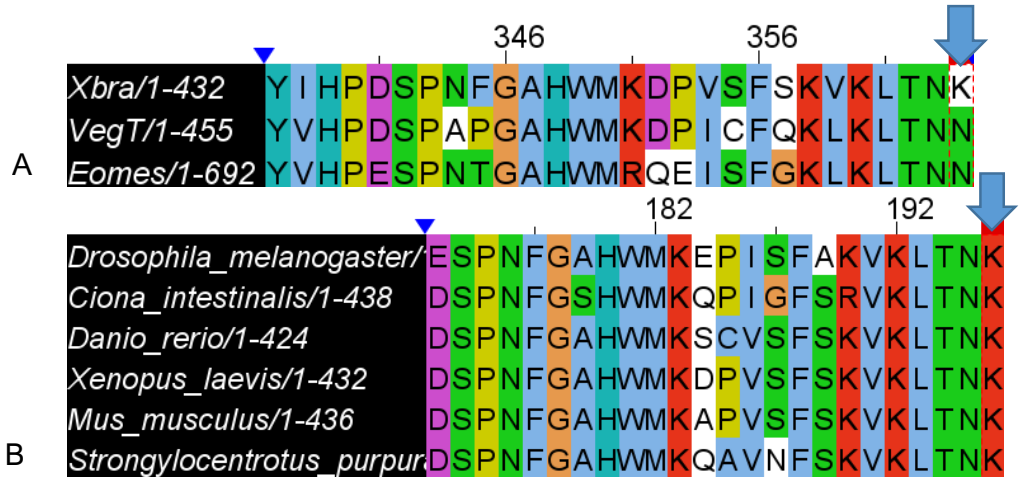


Figure 1.2 Representation of amino acid residues involved in binding to the DNA molecule. A) The residue K (lysine) represents a crucial difference between other T-box DNA-binding residues in *Xenopus* (Xbra, Eomes, VegT) B) Different species' Brachyury sequence alignments and the presence of the residue K (fruit fly, ascidian, zebrafish, frog, mouse, sea urchin). (Adapted and modified from Papaioannou, 2014)

However, the hypothesis based on the lysine presence seems to be only partially correct. The binding specificity and the behavior, such as being activators or repressors of different genes, could lay in protein-protein interactions and various signaling events (Messenger *et al.*, 2005; Marcellini, 2006; Faial *et al.*, 2015). It has been proven that Brachyury co-operates through its N-terminal domain with the MH2 domain of a cofactor Smad1, while Eomes co-operates with Smad2/3 (Messenger *et al.*, 2005; Faial *et al.*, 2015). Moreover, it seems that Brachyury and Eomes have many mutual binding sites, but their regulatory role depends on the signaling they receive, which determines whether they can act as activators or as repressors. A differentiating human embryonic stem cell genome-wide study showed that Brachyury is able to activate mesodermal genes when present in an embryonic region that receives BMP signaling (posterior primitive streak or mesoderm progenitors) and to repress the endodermal genes when present in the region that receives Activin A/Nodal signaling (anterior primitive streak or endoderm progenitors), accompanied by the interaction with Smad1 and Smad2/3/Eomes, respectively (Faial *et al.*, 2015). In addition, Brachyury seems to have the ability to

repress endodermal genes in the posterior primitive streak when it is not co-operating with Smad1 (Faial *et al.*, 2015). Eomes' ability to induce endodermal genes in a synergic manner with Brachyury and to repress mesodermal genes is under the influence of Nodal signaling in the anterior primitive streak (Faial *et al.*, 2015). This is represented in Figure 1.3. Interaction of Brachyury with Smad proteins is discussed in more detail in Chapter 3 and Chapter 6.

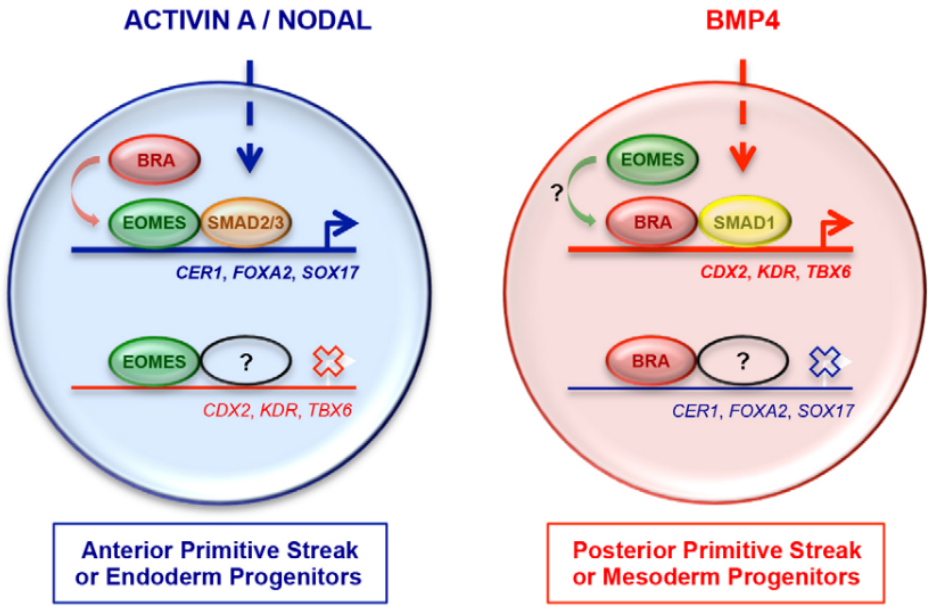


Figure 1.3 Synergistic activity of Brachyury, Eomesodermin and Smad proteins in common domains of expression during human embryo development. Activin A/Nodal signaling is upstream activator of Smad2/3, while BMP4 is the upstream activator of Smad1. (Adapted from Faial *et al.*, 2015)

Furthermore, Wilson and Conlon demonstrated that Brachyury binds exclusively to two core motifs arranged head-to-head, while other T-Box members, such as VegT, bind solely to two core motifs arranged tail-to-tail (Wilson and Conlon, 2002).

T-box binding motifs are highly conserved among different organisms and different members of the T-box protein family. For instance, *Capsaspora* Bra, mouse Bra, and different T-box proteins coming from the same species (mouse Bra, Eomes, Tbx1, Tbx2, and Tbx4) show conserved core motif, as shown in Figure 1.4 (Sebé-Pedrós and Ruiz-Trillo, 2017).

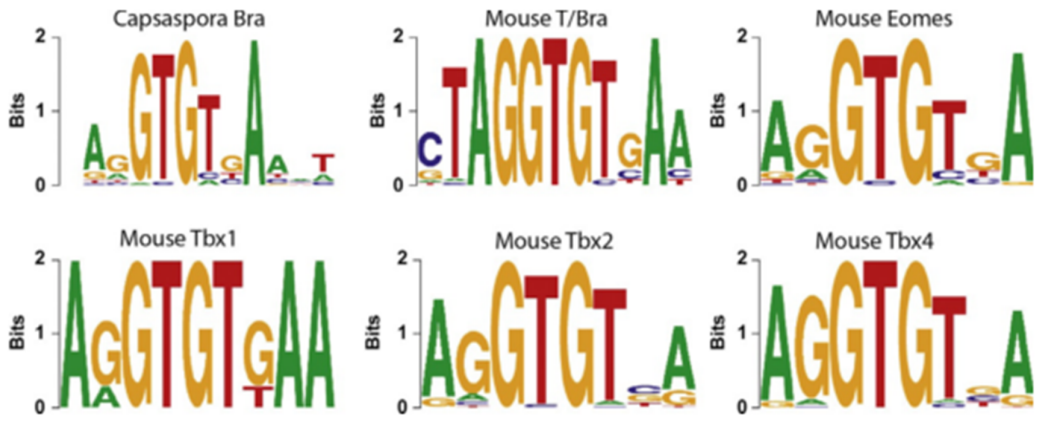


Figure 1.4 Representation of highly conserved T-box binding motifs in different classes of T-box proteins coming from different animals. Adapted and modified from (Sebé-Pedrós and Ruiz-Trillo, 2017) and (Sebé-Pedrós, Ariza-Cosano, *et al.*, 2013)

When first discovered, *brachyury* function was described as being involved in mesoderm differentiation and notochord formation, and further studies on mice had demonstrated this hypothesis being true (Herrmann *et al.*, 1990; Wilkinson, Bhatt and Herrmann, 1990, 1997; Evans *et al.*, 2012). Later on, several groups have isolated *brachyury* gene from zebrafish (*Danio rerio*) and frog (*Xenopus laevis*) developing embryos, where this function was proven to be conserved among vertebrates (Schulte-Merker *et al.*, 1992; Kispert and Herrmann, 1993; Conlon *et al.*, 1996; Evren *et al.*, 2014).

Two independent studies showed consistency with *brachyury's* conserved role in vertebrates and other chordate groups: cephalochordates and tunicates (urochordates). Cloning and testing *brachyury* homolog in *Amphioxus* showed its expression in mesodermal tissue and notochord during development (Holland *et al.*, 1995; Terazawa and Satoh, 1995; Onai *et al.*, 2009). Expression of *brachyury* during tunicate development was studied in two species belonging to ascidians, *Halocynthia roretzi* and *Ciona intestinalis* (*C. robusta*), and one study in larvaceans (*Oikopleura dioca*). The expression pattern was consistent with the vertebrate *brachyury* expression in the notochord (Nakatani *et al.*, 1996; Bassham and Postlethwait, 2000; Katikala *et al.*, 2013).

The closest sister group to chordates, ambulacrarians, which include echinoderms and hemichordates, shows similar *brachyury* expression patterns, different from chordates. In hemichordate, *Ptychodera flava brachyury* expression was detected in gastrulating embryos in oral and anal regions. In the larval stages, the expression is abolished. In contrast, during metamorphosis, *brachyury* expression reappears in the mesoderm of the protosome, mesoderm of the collar, and trunk, in the proboscis musculature, posterior region of the metacoelomic somatopleura, and the wall of the gut. Later, by 72 hours, juveniles show *brachyury* in the mesoderm of all three body regions and the posterior gut. The mesodermal expression could be homologous to the notochord expression in chordates (Peterson, Cameron, *et al.*, 1999).

In echinoderms, the expression of Brachyury was studied during the early development in sea urchins: *Hemicentrotus pulcherrimus* (Harada, Yasuo and Satoh, 1995; Hibino *et al.*, 2004), *Paracentrotus lividus* (Croce, Lhomond and Gache, 2001), *Lytechinus variegatus* (Gross and McClay, 2001), *Strongylocentrotus purpuratus* (Peterson, Harada, *et al.*, 1999; Rast *et al.*, 2002), starfish: *Asterina pectinifera* (Shoguchi, Satoh and Maruyama, 1999) and sand dollars: *Clypeaster japonicus*, *Astriclypeus manni*, *Peronella japonica* and *Scaphecinus mirabilis* (Hibino *et al.*, 2004). *Brachyury* transcripts were detected as early as in the blastula stage at the mid-blastula's vegetal pole in the ring of the presumptive endoderm in the cells called veg2 (Figure 1.5A). This transcription factor is dynamically expressed in all veg2 endodermal cells before entering the archenteron and becoming part of the gut (Figure 1.5B). When veg1 cells are recruited as the new endodermal lineage, *brachyury* continues to be expressed in the ring of cells surrounding the blastopore (Figure 1.5C). Another expression domain is the oral ectoderm that appears in the late mesenchyme blastula, which will give rise to the mouth in the pluteus larva (Figure 1.5B, 1.5C). At the larval stage, *brachyury* is expressed in the mouth and anus of pluteus larva in sea urchins and bipinnaria larva of starfish (Peterson, Harada, *et al.*, 1999; Shoguchi, Satoh and Maruyama, 1999; Croce, Lhomond and Gache, 2001; Gross and McClay, 2001; Rast *et al.*, 2002; Hibino *et al.*, 2004). In sand dollars, the expression of *brachyury* follows the same pattern as in sea urchins and starfish. However, two species *A. manni* and *S. mirabilis*, showed an additional third

domain of expression – the oral side of the archenteron in the mid-gastrula and early prism stage, respectively (Hibino *et al.*, 2004).

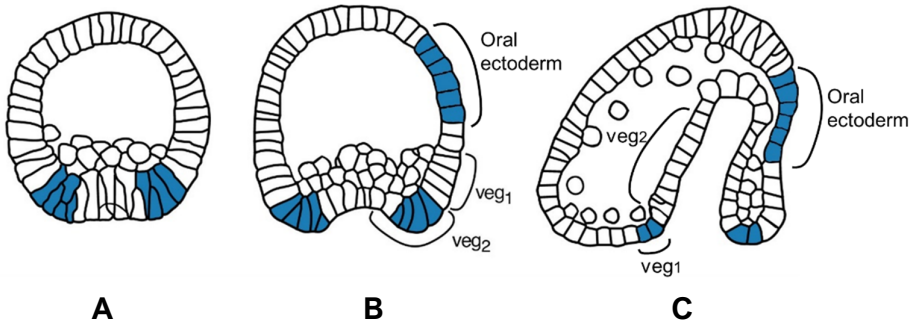


Figure 1.5 Schematic representation of the *brachyury* expression in the sea urchin embryo. The territories of expression are shown in blue. (a) Early mesenchyme blastula (b) Early gastrula (c) Late gastrula

The expression of *brachyury* was studied in another major bilaterian group, the Protostomia, in order to explore its evolution and ancestral role. In all major protostome groups studied so far, apart from Platyhelminthes and Nematoda, the *brachyury* gene appears to be present in the genome (Sebé-Pedrós and Ruiz-Trillo, 2017). In Nematoda, such as *Caenorhabditis elegans*, there seem to be 19 T-box genes, and one of them, *mab-9,* shows similar function to *brachyury*, being involved in hindgut development (Woollard and Hodgkin, 2000).

In Arthropoda, *brachyury* expression was assessed in four insect species, in the fruit fly *Drosophila melanogaster*, in the beetle *Tribolium castaneum*, in the grasshopper *Locusta migratoria*, and in the cricket *Gryllus bimaculatus* (Kispert *et al.*, 1994; Lengyel and Iwaki, 2002; Shinmyo *et al.*, 2006; Berns *et al.*, 2008). Even though these insect species belong to different orders and show differences in embryonic, larval, and adult morphologies, and go through different metamorphosis types until they reach their adult forms, *brachyury* expression is conserved during the early development. Its expression initiates in the ring of cells that will eventually internalize to form the gut and continues its expression in the hindgut during the embryogenesis (Kispert *et al.*, 1994; Lengyel and Iwaki, 2002; Shinmyo *et al.*, 2006; Berns *et al.*, 2008; Hejnol and Martín-Durán, 2015).

In Mollusca, *Brachyury* was studied in the gastropod *Patella vulgata*. In *P. vulgata*, *brachyury* expression starts very early during the first cleavage divisions. It is involved in the anterior-posterior axis establishment while expressed in both mouth and anus (Lartillot *et al.*, 2002).

In Annelida, *brachyury* shares a conserved expression pattern with basal deuterostome and protostome members. In *Hydroides elegans*, expression of *Brachyury* is involved in morphogenetic events associated with gastrulation by invagination in the endoderm and morphogenetic events in the ectoderm. Its expression is initiated quite early in the blastomeres that drive gastrulation (Arenas-Mena, 2013). In *Platynereis dumerilii*, the expression starts in vegetal cells around the closing blastopore and later during development in the mouth and anal region of a developing trochophore larva. Some visceral mesoderm cells show additional *brachyury* domain of expression. Later during development, *brachyury* expression reappears in the ventral midgut of the young worm (Arendt, Technau and Wittbrodt, 2001).

Brachyury orthologues were also found in non-bilaterian groups: Cnidaria, Placozoa, Ctenophora, and Porifera.

In the sea anemone, *Nematostella vectensis* (Anthozoa), the first signs of *brachyury* expression are at the late blastula/early gastrula stage in the blastopore ring, the expression is restricted to the same region during the planula stage, later it extends to the mesenteries during the polyp stage, and the expression is continuous into adulthood (Scholz and Technau, 2003). *Brachyury* knock-out experiments in *N. vectensis* showed that this gene is essential for early development and that it is involved in endoderm development, pharynx formation, and oral-aboral patterning (Servetnick *et al.*, 2017).

In two different species of Anthozoa, the corals *Acropora millepora* and *Acropora digitifera Brachyury* expression in the pre-gastrula stage (prawn-chip stage) showed that this gene potentially might be the critical gene in setting the boundary between ectoderm and endoderm, which is not the case in *N. vectensis* (Hayward *et al.*, 2015; Yasuoka, Shinzato and Satoh, 2016).

The hydrozoan *Clytia hemisphaerica* has the conserved cnidarian *brachyury* expression in the oral ectoderm during the early development (Momose, Derelle and Houliston, 2008). The analysis of genes involved in Wnt-dependent embryo patterning demonstrated that orally expressed *brachyury* is a duplicated gene. Knock-down experiments of Brachyury showed that *brachyury* paralogues are both involved in morphogenetic movements upstream of endoderm specification during the early development (Lapébie *et al.*, 2014). Both paralogues *Bra1* and *Bra2* are expressed in the embryo's oral pole (blastopore does not form during *Clytia* development), and their absence causes a delay in the gastrulation (Lapébie *et al.*, 2014). They have distinct roles in ectoderm and endoderm layers in the polyp stage, *Bra1* being predominantly endodermal, and *Bra2* being mostly expressed in the ectoderm (Bielen *et al.*, 2007).

Brachyury function has been studied only in two Porifera (Sponges) species, in demosponge *Suberites domuncula* and calcarean sponge *Sycon ciliatum* (Adell *et al.*, 2003; Leininger *et al.*, 2014). In *S. domuncula*, after the dissociation highest expression levels were in aggregates of dissociated cells, which implies its function in cell adhesion, morphogenetic movements, and cell motility (Adell *et al.*, 2003). In *S. ciliatum* two *brachyury* paralogues were found, expressed in different cell types: oocytes, in embryonic micromeres, and adult choanocytes, that make up the choanoderm, which could imply the conserved role homologous to forming eumetazoan endomesoderm (Leininger *et al.*, 2014).

In Placozoa, *brachyury* is expressed in a few cells or groups of cells in smaller and potential outgrowth zones in larger animals (Martinelli and Spring, 2003).

Brachyury is present in the genome of a ctenophore *Mnemiopsis leidyi*. The localization of expression is around the blastopore and in the invaginating stomodeum. The most probable function of *Brachyury* in the development of *M. leidiy* is controlling morphogenetic movements during gastrulation and the formation of stomodeum and pharynx (Yamada *et al.*, 2010).

Finally, there is an interesting question if there is a conserved *brachyury* function between non-metazoans and metazoans. The unicellular species *Capsaspora owczarzaki* (class Filasterea) and chytrid fungus *Spizellomyces punctatus* contain

brachyury gene in their genomes (Sebé-Pedrós *et al.*, 2011). *C. owczarzaki* is in the key phylogenetic position for studying the evolution of multicellular organisms. It is an amoeboid holozoan whose life cycle includes the stage of multicellular aggregations (Sebé-Pedrós *et al.*, 2011; Sebé-Pedrós, Irimia, *et al.*, 2013; Booth and King, 2016). Studying its transcriptional regulation is crucial for understanding the evolution of metazoans. These studies showed that vital genetic machinery and signaling mechanisms normally present in metazoans are also existing in *C. owczarzaki*, including the Brachyury transcription factor (Sebé-Pedrós *et al.*, 2011, 2016; Suga *et al.*, 2013; Booth and King, 2016).

It seems that the *brachyury* has a conserved role in morphogenesis in bilaterian and non-bilaterian metazoan taxa (Figure 1.6). Its function is in different developmental aspects and cell-type specifications. Those will be described further in this chapter. This overview shows that *brachyury* has very early origins in evolutionary history and that its function in cell motility might be conserved among most of the metazoans and holozoans (Sebé-Pedrós *et al.*, 2011, 2016; Suga *et al.*, 2013; Booth and King, 2016).

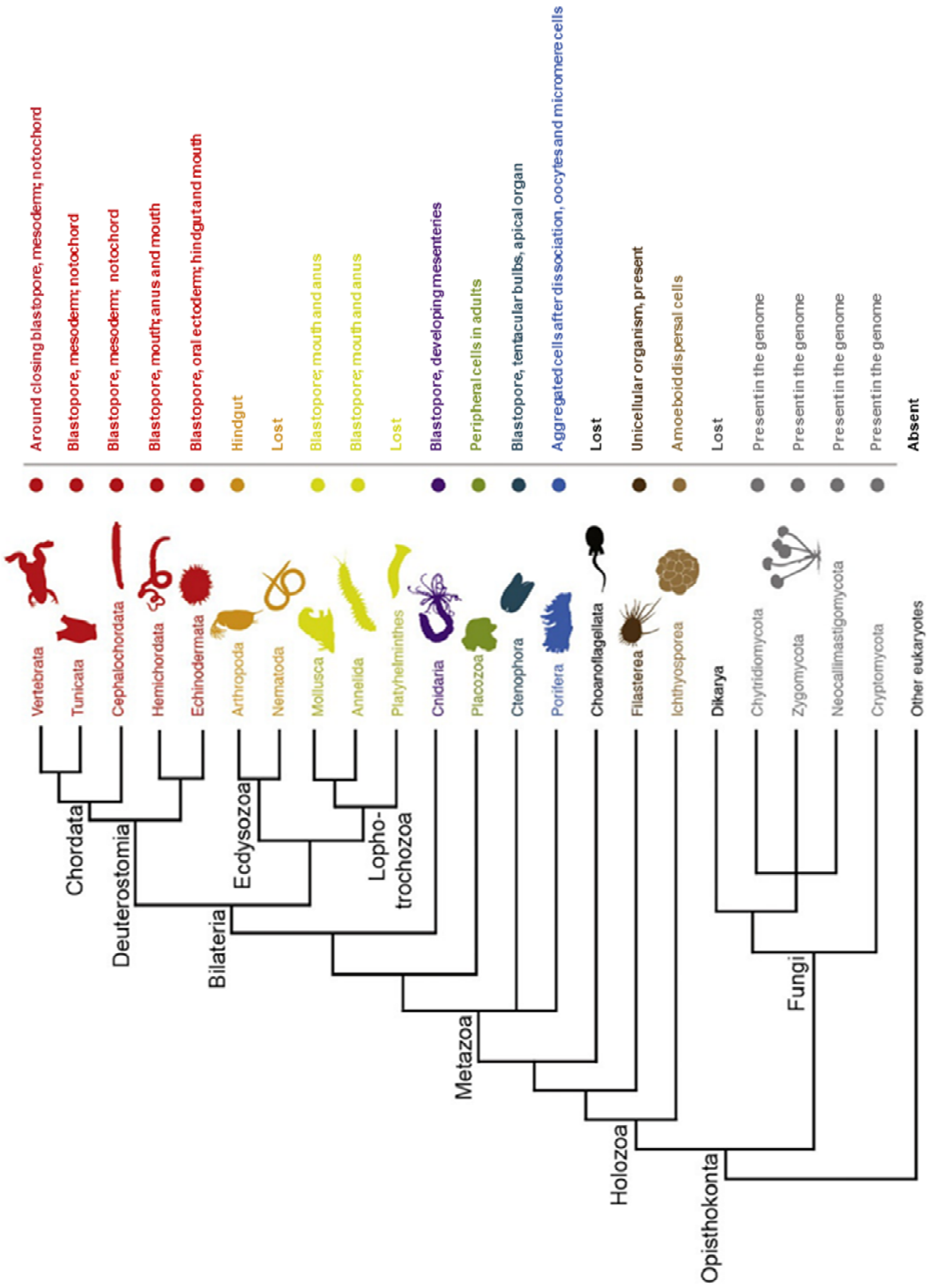


Figure 1.6 Expression domains of Brachyury transcription factor among different taxa. Colored dots represent the presence of Brachyury. The column after the dots shows spatial expression pattern of *brachyury* earlier and later during development. All metazoans contain at least one copy of the *brachyury* gene. Nematodes and platyhelminths had lost *brachyury* during their evolution. In non-metazoan organisms, *brachyury* is present in Filasterea, Ichthyosporea and Fungi. Metazoan sister group Choanoflagellata does not have *brachyury* gene. (Adapted and modified from Sebé-Pedrós and Ruiz-Trillo, 2017)

How did Brachyury's function change from its primitive role in cellular motility in unicellular organisms to the most derived and crucial role in forming the notochord in chordates? It seems that in the basal Metazoa Brachyury was co-opted for being involved in the gastrulation, and particularly in the germ layer segregation and differentiation. Moving up the phylogenetic tree, we see that Brachyury gets more and more involved in the regulation of the tissues forming around the blastopore: future mouth and anus in protostomes and future anus and mouth in deuterostomes. In chordates, Brachyury's role in regulating posterior gut diminished, even though the expression is seen around the blastopore at the onset of gastrulation. To detangle what could potentially be the transitional role of Brachyury, one should look at the chordates' sister group, Ambulacraria. Ambulacraria, still a deuterostome group, is composed of Hemichordata and Echinodermata. This group is phylogenetically closer to chordates, but it seems that the role of Brachyury is similar to most protostome groups. Echinoderms are basal deuterostomes and the most important link between the chordates and protostomes. They are an indispensable link in understanding chordate origins and separating chordate genetic and morphological traits.

1.3 Sea urchins as model organisms in evo-devo studies

Sea urchins are Ambulacraran Deuterostomes belonging to the phylum Echinodermata and the class Echinoidea. The name "Echinodermata" comes from them having spiny skin (in Greek, "ekhinos" means "spiny," and "dermos" means "skin"). They have radial symmetry as adults, while during larval stages, they exhibit bilateral symmetry (Sprinkle, 1992). Sea urchins are convenient model organisms for developmental biology studies because of their relatively short development, numerous gametes produced (around 50 000 000 eggs in one spawning), external high-rate fertilization, synchronous development, easy culturing of embryos and larvae, and their transparent morphology (McClay, 2011). Many molecular tools are available; for example, the gene function in these organisms is simple to perturb and manipulate. DNA and mRNA microinjections are efficiently performed into the eggs,

and DNA incorporates into the blastomere nucleus around the second, third, or the fourth cleavage stages (Arnone, Dmochowski and Gache, 2004; Cheers and Ettensohn, 2004; Lepage and Gache, 2004; McClay, 2011). Another significant benefit of working with sea urchins is that the whole genome of *Strongylocentrotus purpuratus* has been sequenced and published (Sodergren *et al.*, 2006). Moreover, the genome of the Mediterranean sea urchin *Paracentrotus lividus* has been recently sequenced; however, it has not been published yet. The genomes of the sea urchins *Allocentrotus fragilis*, *Hemicentrotus pulcherrimus*, *Strongylocentrotus franciscanus*, *Lytechinus variegatus*, *Arbacia punctulata*, *Eucidaris tribuloides*, the sea star *Patiria miniata*, the sea cucumber *Parastichopus parvamensis*, and the brittle star *Ophiothrix spiculata* are partially available (echinobase.org), making echinoderms very convenient to work within developmental and molecular studies, specifically gene regulatory networks (GRNs) studies (Arnone, Byrne and Martinez, 2015). Sea urchins are good models for evolutionary studies due to their phylogenetic position. Echinoderm phylum diverged during the Cambrian period (Figure 1.6) 500-540 million years ago (McClay, 2011). Their position in Deuterostomia makes them an excellent experimental taxon for evolutionary developmental biology (evo-devo) studies because they are with their sister group, Hemichordates, closest relatives of Chordates.

This project's model systems are two sea urchin species: the Pacific species *Strongylocentrotus purpuratus* and the Mediterranean one *Paracentrotus lividus*. Even though they both belong to the order Echinoidea, their evolutionary distance is 40 Mya, as shown in Figure 1.7. *P. lividus* was selected as a comparison species thanks to its evolutionary distance to *S. purpuratus* and also because it is easily available for collection in the Gulf of Naples, but most importantly due to its recently assembled genome.

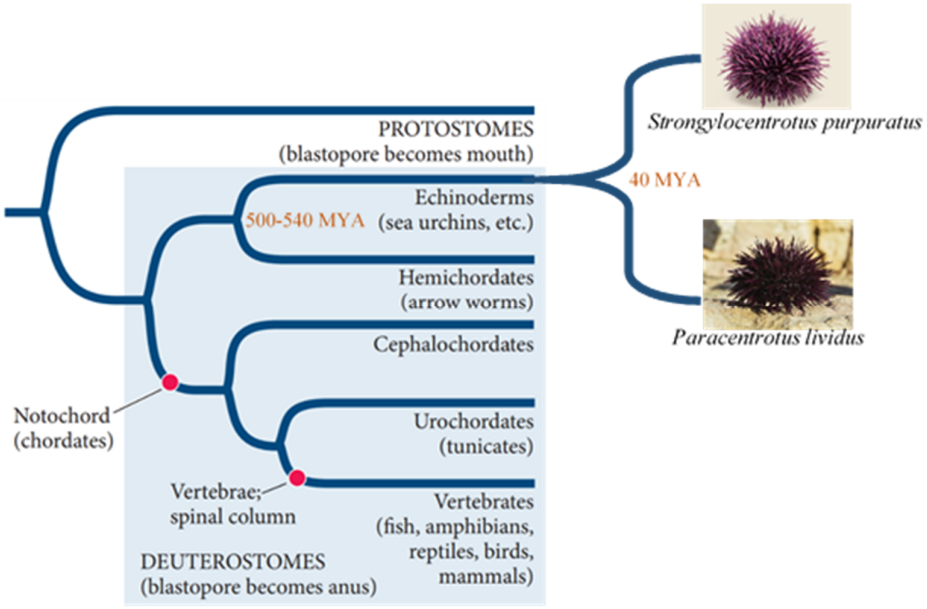


Figure 1.7 Phylogenetic position of model systems used in this project. (Adapted and modified from Gilbert and Barresi, 2017)

Sea urchins have been used for developmental studies for decades now, and their gastrulation process became a model for deuterostome morphogenetic process studies due to their transparent morphology and fast development.

The sea urchin fertilization starts with the fusion of gametes and, subsequently, the fusion of the male and female nuclei in a pronuclear fusion process. After fertilization, the fertilization envelope rises, protecting the egg from polyspermy. The zygote has two territories, the animal and the vegetal pole (Figure 1.8A). Soon after the fertilization, the zygote starts to divide. The first three cleavages divide the fertilized egg into cells of the same size (Figure1.8A-D). The fourth cleavage is different. The animal part divides into eight equally sized cells termed as mesomeres, and the vegetal part divides unequally, and the two types of cells are produced – the upper, four larger macromeres, and the lower, two smaller micromeres (Figure 1.8E). The fifth cleavage will give rise to the equally divided mesomeres forming an1 tier and an2 tier, as well as the macromeres that divide into eight cells (Figure 1.8F). However, the micromeres divide unequally, giving rise to four large and four small micromeres (Figure 1.8F). After the sixth cleavage, the 60-cell embryo is formed.

The 60-cell embryo animal pole is composed of two layers of 8 an1, and 8 an2 cells placed one above the other. The vegetal pole is composed of veg1 tier of cells, veg2 tier of cells, and 12 micromeres: 8 large and 4 small (Figure 1.8G). Before the last cleavage division, the sea urchin embryo has already specified cell territories – the future vegetal plate territory (that will give rise to endoderm and non-skeletogenic mesoderm during the gastrulation process), oral and aboral ectoderm, skeletogenic mesenchyme, and small micromeres. The last cleavage will give rise to a 128-cell embryo, also known as blastula (Figure 1.8H) and the cell division continues, making the cells progressively smaller and smaller while the embryo stays the same size.

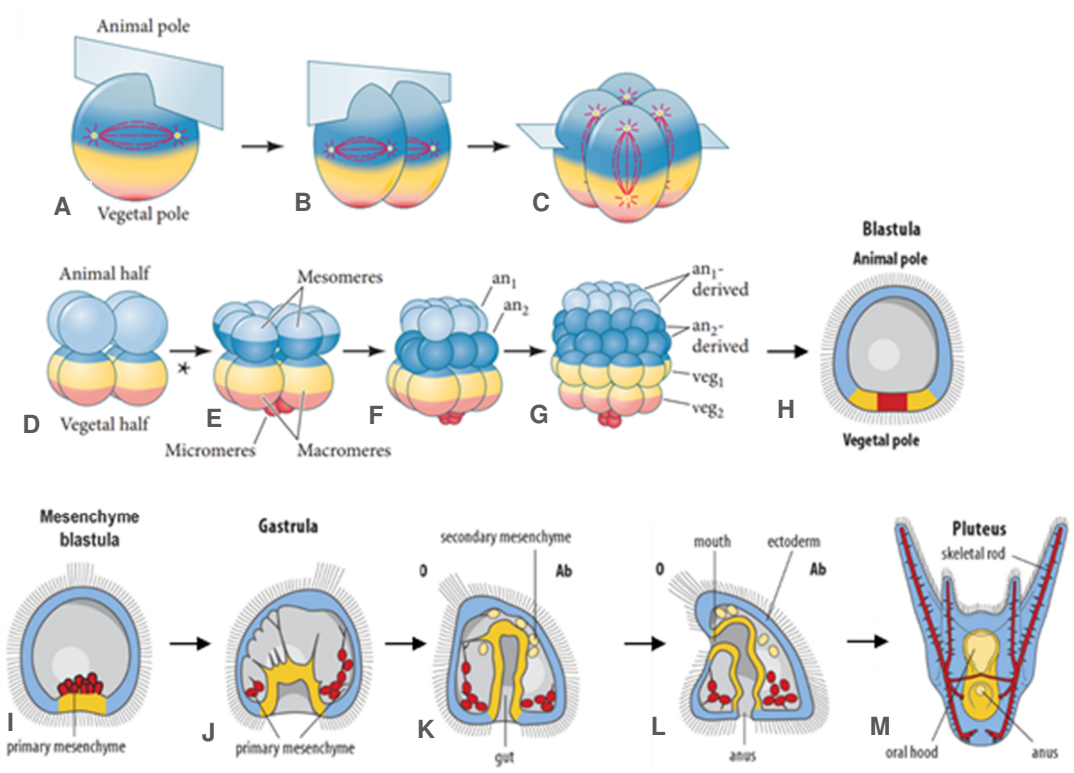


Figure 1.8 Early development of the sea urchin. A) Zygote undergoing the first cleavage; B) Two blastomeres undergoing the second cleavage; C) Four blastomeres undergoing the third cleavage; D) 8-cell developmental stage undergoing the fourth cleavage; E) The embryo is composed of 8 animal half mesomeres, 4 vegetally localized macromeres and 2 micromeres; F) The fifth cleavage is completed and embryo is made up from an1 and an2 mesomere tier, eight macromeres and four micromeres; G) The sixth cleavage is completed giving rise to 60-cell stage embryo; H) 128-cell stage or blastula; I) Mesenchyme blastula with ingressed primary mesenchyme cells (PMC); J) Mid-gastrula stage with invaginating gut and moving PMC cells; K) Late gastrula with forming secondary mesenchyme cells (SMC); L) Prism-stage embryo with formed mouth, anus and developing skeleton; M) Pluteus larva with completely formed digestive tract and skeletal rods. Ectodermal cells are in colored blue, mesodermal cells, and derivatives in red and endodermal cells in yellow. O – oral, Ab – aboral (Adapted and modified by Gilbert and Barresi, 2017; Green and Batterman, 2017)

Blastula has a shape of a hollow ball surrounding the internal cavity termed the blastocoel. The vegetal cells start to thicken, giving rise to a vegetal plate. The animal pole will develop a tuft of cilia that will allow the embryo to swim after hatching (Figure 1.8I). The skeletogenic mesenchyme cells, also known as primary mesenchyme cells (PMCs), start to ingress and fill the blastocoel adjacent to the vegetal plate. This developmental stage is known as mesenchyme blastula (Figure

1.8J). PMCs begin to move and re-arrange themselves in the form of a ring. The PMC ring surrounds the invaginating veg2 tier of cells. As veg2 cells invaginate, they start to form an opening known as the blastopore (future anus). The veg2 lineage, specified already before the last cleavage, give rise to the veg2 endodermal linage that will form a gut and secondary mesenchyme cells (SMC). The four small micromeres stay on top of the veg2 meso lineage, and they do not divide. Veg2 endodermal cells proliferate and move upward, forming a primitive gut (archenteron) in the process known as gastrulation by being pushed from the epithelial cells adjacent to the vegetal plate. This leads to the mid-gastrula stage, and this process is known as the primary invagination that forms a tube-like archenteron (Figure 1.8K). As the veg2 endodermal cells invaginate, the second endodermal lineage or veg1 endodermal cells move towards the center of the vegetal plate surrounding the blastopore. The SMCs start to extend their filopodia towards the apical plate, resulting in pulling the archenteron upwards and extending it further. While the gut is growing, SMCs begin to proliferate, and they give rise to four types of non-skeletogenic mesoderm cells (NSM), which are known as the blastocoeliar cells, pigment cells, left and right coelomic pouch cells, and circumesophageal muscle cells (Figure 1.8K). At the end of gastrulation, the gut bends towards the oral ectoderm region, where it will fuse with it, generating another opening – the future mouth. At this time skeleton is partially formed by the PMCs, and the embryo takes the prismatic shape; hence this stage is known as the prism stage (Figure 1.8L). At this stage, the gut starts to take the form of a tripartite structure, parts of which will later be known as the foregut, midgut (stomach), and the hindgut in the larval (pluteus) stage (Figure 1.8M) (Arnone, Byrne and Martinez, 2015).

1.4 Gene Regulatory Networks for development and evolution

Gene regulatory networks (GRNs) represent a combination of interactions between the regulatory genes, transcription factors, and signaling molecules that establish specific regulatory states in the spatial and temporal domain of the developing organism (Davidson, 2006; Peter, 2017).

Regulatory states are the sets of co-expressed transcription factors at a specific time and space, and they determine the integrative control of gene expression (Peter, 2017). GRNs describe post-embryonic developmental processes – organogenesis, the formation of body parts, cell-type specification, and physiological capabilities. GRNs contain both cis and trans-regulatory elements. Cis-regulatory elements represent DNA sequences that regulate the corresponding genes' spatial and temporal expression depending on the information they receive from trans-regulatory elements, transcription factors that are active in each cell at a specific time of development (Erwin and Davidson, 2009; Li and Davidson, 2009). Transcription factors (TFs) are proteins that can recognize and bind specific short DNA sequences in the above-defined cis-regulatory regions of a given gene and control its expression, either by activating it or suppressing it (Latchman, 1997; Peter and Davidson, 2015). Transcription factors are composed of DNA-binding domains by which they are defined and classified in different TF families and effector domains that cause interactions with other transcription necessary proteins and other transcription factors. The DNA-binding domains show significant evolutionary conservation, whereas the effector domains evolve at a greater rate (Peter and Davidson, 2015). TFs always act together with other transcription factors and cofactors (Thomas and Chiang, 2006; Sikorski and Buratowski, 2009; Peter and Davidson, 2015).

Cis-regulatory modules (CRMs) represent DNA sequence elements that have regulatory activity on transcription. They can be promoter sequences that bind the transcription machinery (RNA polymerase II and cofactors); they can be insulators that divide regulatory domains one from another; they can be enhancers that activate gene transcription by the binding of TFs that act as activators, or they can be silencers that suppress gene expression by binding of TFs that act as repressors (Serfling, Jasin and Schaffner, 1985; Gaszner and Felsenfeld, 2006; Maeda and Karch, 2011; Kolovos *et al.*, 2012; Tippens, Vihervaara and Lis, 2018). Transcription factor binding sites can be close or further away from the transcription start sites; they can exist randomly in the genome every few thousand base pairs; however, the regulatory specificity is determined by being tightly clustered with other TF binding sites (Peter and Davidson, 2015).

Even though GRNs are highly conserved, they can experience evolutionary changes over time (Peter and Davidson, 2011b; Erkenbrack, Davidson and Peter, 2018). GRN evolutionary changes can occur in both levels of cis-regulatory and trans-regulatory elements. The changes in cis-elements affect regulatory sequences which control the expression of the TF itself, if this element is close to the TF coding gene, or changes in the TF DNA-binding sites in the genome (Davidson and Levine, 2005; Davidson, 2006; Erwin and Davidson, 2009; Peter and Davidson, 2011b, 2015; Sorrells and Johnson, 2015). The changes in trans level could represent changes in cofactor interaction (Marcellini *et al.*, 2003; Marcellini, 2006), the changes in the coding sequence of the transcription factor that can influence its protein structure (Starr and Thornton, 2016) and DNA-binding specificities of a TF (Hudson et al., 2015).

GRN evolutionary changes in both cis-regulatory and trans-regulatory elements lead to TFs getting new downstream targets and to the TF network rewiring (Sebé-Pedrós and Ruiz-Trillo, 2017). These changes are of great importance since they can affect the evolution of specific body parts and body plans overall. Evolutionary changes of transcription factors can have two main consequences:

1. Loss of their original function (loss of the ability to bind to the original cis-regulatory element)

2. Gain of a new function (ability to recognize and bind to a new cis-regulatory region).

Both consequences could lead to the rewiring of the original GRNs.

In conclusion, GRNs represent a conceptual map that puts together cis-trans interactions occurring during cell type specification or developmental processes at a given time and space (Erwin and Davidson, 2009; Li and Davidson, 2009).

In this thesis, I will discuss the rewiring of the GRNs downstream of the Brachyury transcription factor during the early development, specifically during the transition from early to late gastrula, in two sea urchin species.

1.5 Aims of the thesis

This project aims to study the evolution of the GRN downstream of Brachyury in two sea urchin species, *Strongylocentrotus purpuratus* and *Paracentrotus lividus*. The combination of different high-throughput omics approaches to study the transcriptional regulation was adopted to reconstruct, extend, and improve the known regulatory networks downstream of Brachyury. This is a comparative study of Brachyury's role in regulating cell-type specification and gut patterning during the early embryonic development of two sea urchin species. The primary goal is to discover putative direct and indirect Brachyury targets by differential transcriptomic approach, by analyzing normal and Brachyury knock-down embryos. The secondary goal is to combine these discoveries with the knowledge of chromatin organization and its accessibility and sensitivity to Brachyury TF binding using the ATAC-seq and Brachyury ChIP-seq data to discover the direct targets of this TF. This work shows the conservation of GRNs around Brachyury and putative novel functions caused by rewiring the network in two sea urchin species. The final goal is to compare Brachyury's role in echinoderm development to the known role in the development of other bilaterian and non-bilaterian groups.

CHAPTER 2

MATERIALS AND METHODS

This chapter contains information about the experimental procedures used and the data analyses performed.

2.1 Animal handling and culturing of sea urchin embryos

Paracentrotus lividus adult animals were obtained from the Gulf of Naples and provided by the marine service of Stazione Zoologica Anton Dohrn. They were kept in large circulating seawater tanks at 18°C and fed with algae. *Strongylocentrotus purpuratus* adult animals were obtained from Kerchoff Marine Laboratory, California Institute of Technology, USA. They were kept at Stazione Zoologica in circulating seawater tanks at 15-16°C and were fed with algae. 37.8 ppt salinity water was used for *P. lividus* and 34.02 ppt for *S. purpuratus*.

2.1.1 Gamete collection

P. lividus adults' spawning was accomplished by intracoelomic injection of 0.5 M potassium chloride (KCl). To preserve the animals, in the case of *S. purpuratus*, spawning was mostly caused by vigorous shaking until the gametes are released. After the injection, or shaking, a female animal was placed onto a 50 ml glass beaker filled with cold filtered seawater (FSW), put on ice, with the aboral side facing the beaker. After the eggs were shed, they were filtered through a 200 μm filter and washed with FSW to remove the debris and coelomic fluid. After the release of gametes, dry sperm is collected from a male animal with a glass Pasteur pipette into a 1.5 ml Eppendorf tube and kept on ice until use.

2.1.2 Culture

Before fertilization, 5 µl of dry sperm was activated in 13 ml of FSW. The fertilization process was detected by the presence of an elevated vitelline membrane. After fertilization, eggs were washed from the excess sperm to avoid polyspermy and placed in clean glassware free of any detergents. Cultures were kept in 5 L glass beakers containing 3 L of FSW to prevent over-crowding. For *P. lividus*, local seawater filtered through a 0.22 µm Millipore was used, and culture was kept in an incubator set to 18°C with the 12-hour day and 12-hour night cycle. *S. purpuratus* cultures were maintained in local seawater diluted 9:1 with distilled water, filtered through 0.22 µm Millipore and kept in an incubator at 15°C with the day and night cycle. Since one of the desired stages was late gastrula, cultures were kept for 24 hours in the case of *P. lividus* and 48 hours in the case of *S. purpuratus* without the addition of antibiotics or food.

2.2 Material preparation and microinjection of morpholino antisense oligonucleotides

2.2.1 Material preparation

Protamine coated plate preparation. The microinjection was performed on the protamine sulfate coated plates, which keep the eggs adhered to the plate bottom. Protamine plates were prepared by pouring 1% protamine sulfate (in dH2O) onto the lid of 60 mm diameter Petri dishes and waiting for 1 minute. After 1 minute, protamine sulfate was removed, and plates were washed several times with dH2O to remove any unadhered protamine sulfate solution and left to air-dry. When dry, protamine plates were scratched in the middle using the razor blade and stored at room temperature.

Preparation of “rowing” glass pipettes. The “rowing” glass pipettes are controlled by mouth and are used to transfer the eggs onto the protamine plates before microinjection. The “rowing” glass Pasteur pipettes were pulled out in a Bunsen

burner flame and broken off at the end. The internal diameter of a pipette was about 80 µm, the same diameter as the egg sea urchin egg.

Preparation of microinjection needles. The needles for microinjection were prepared from borosilicate glass supplied by Sutter Instrument Co. Novato, CA (No. B100-75-10) with 1.0 mm outside diameter and 0.75 mm inside diameter. Fine-tipped microinjection needles were pulled on a Sutter P-97 micropipette puller (P=300; H=560; Pu=140; V=80; T=200). Before injection, needles were touched against the scratch on the microinjection dish to break open the needle's tip to a diameter of ~0.4-0.9 µm.

Preparation of solutions for microinjection. All solutions to be microinjected have been filtered through 0.22µm PVDF micro-filter (Millipore) and centrifuged for at least 15 minutes in a micro-centrifuge at maximum speed.

Sea urchin egg preparation. Prior to microinjection, the eggs were de-jellied. In the case of *S. purpuratus*, the eggs were washed in FSW (9:1) and passed a couple of times through a 200 µm Nitex filter. *P. lividus* eggs preparation included an extra step of treatment in acidic seawater (75 µl of 1M citric acid in 50 ml of FSW, pH of 4.2) on agarose plate by swirling (60 mm plastic Petri-dish bottom containing 1% agarose diluted in FSW) until the eggs start to concentrate in the middle and touch each other. After treatment in acidic seawater, the eggs were transferred to 1% agarose plates containing FSW.

After being de-jellied, using the “rowing” pipette, the eggs were transferred in groups of hundreds in a line to the protamine plates in 10 ml of FSW containing 2mM para-aminobenzoic acid (PABA; PS-SW + PABA). The eggs were fertilized directly on the protamine plates by adding 2-3 drops of diluted sperm. The microinjection needle was quickly moved, with the solution flowing, in and out of the egg. The procedure was repeated for the next egg in line until the entire plate was injected.

2.2.2 Morpholino Antisense Oligonucletiode mediated gene knock-down

De-jellied fertilized *S. purpuratus* and *P. lividus* eggs were fixed to protamine sulfate coated plastic Petri dishes and injected using a glass micropipette.
The translation blocking morpholino antisense oligonucleotide against *S. purpuratus brachyury* sequence (SpBra MASO) was available in the lab. The sequence used was CGCTCATTGCAGGCATAGTGGCG (Rast *et al.*, 2002). The translation blocking MASO against the *P. lividus brachyury* sequence (PlBra MASO) was newly designed using the sequence CTGGCAGAAGATGTACTTCGACGAT. MASOs were resuspended in dH2O to a stock solution of 500 µM. A working solution of SpBraMASO 200 µm and PlBraMASO 250 µM of morpholino oligos in 0.2 M KCl was injected into fertilized eggs (2 to 4 pl) of *S. purpuratus* and *P. lividus*, respectively.
In all experiments, as a negative control, embryos were injected with 100 µM of the standard control MASO sequence CCTCTTACCTCAGTTACAATTTATA, at equal or greater concentration, and compared together with uninjected and MASO injected embryos. The MASOs were acquired from Gene Tools (Corvallis, OR).
The experiment was performed in three biological replicates for *S. purpuratus* and four biological replicates in *P. lividus* for each stage: the early and late gastrula. Each replicate contained 100-150 fertilized sea urchin eggs. Dr. M.I. Arnone performed the microinjection procedure.

2.3 RNA-seq workflow

2.3.1 RNA extraction, sequencing, quality check, read mapping and quantification

RNA extraction and quality assessment. RNA was extracted using the RNAqueous-Micro Total RNA Isolation Kit (Invitrogen) from 700 embryos and checked for quality using Agilent 2100 Bioanalyzer Instrument. Samples with the RIN number higher or equal to 7 were sent for sequencing. The RIN number represents the ratio between

two ribosomal bands of 28S and 18S subunits, and it is a good estimator of RNA quality and degradation (Schroeder et al., 2006). Periklis Paganos, from Dr. Arnone's lab, has conducted the RNA-extraction. The extracted RNA was retrotranscribed to make cDNA libraries using the Poly-A enrichment approach following the TruSeq protocol, and they were sequenced using the Illumina NextSeq 500 platform. The sequencing depth was from ~11 to ~20 million reads, and 75 base pair (bp) long paired-end reads were generated. The sequencing service was provided by the Laboratory of Molecular Medicine and Genomics at the University of Salerno (http://www.labmedmolge.unisa.it/).

Read quality filtering. To detect any low quality reads, presence of the sequencing adapters or contaminant reads of viral, bacterial, or human origin and, subsequently, remove them, the sequenced reads (fastq files) went through quality check using the default settings of FASTQC v0.11.5 tool (Andrews and Babraham Bioinformatics, 2010). TruSeq sequencing adapters were removed with Trimmomatic v0.38 using the default ILLUMINACLIP setting (Bolger, Lohse and Usadel, 2014). The *Trimmomatic* paired output files were used as input files for *Salmon* v0.11.3 (Patro *et al.*, 2017).

Mapping and read quantification. *S. purpuratus* reads were mapped to the Transcriptome sequence v3.1 (Tu, Cameron and Davidson, 2014), a part of Genome assembly v3.1 available at the www.echinobase.org (Kudtarkar and Cameron, 2017). *P. lividus* reads were mapped to the Transcriptome sequences of the first *P. lividus* Genome assembly that is not published yet (Dr. Arnone's lab is part of sequencing consortium), using the default settings with --gcBias flag in order to correct for fragment-level GC biases in the input data *Salmon* v0.11.3 tool (Patro *et al.*, 2017). For *S. purpuratus Salmon* index was generated using *S. purpuratus* Transcriptome FASTA file from www.echinobase.org and for *P. lividus* using the *P. lividus* Transcriptome FASTA file (unpublished) and the with k value equal to 25. The *Salmon* output files generated count tables for each mapped and quantified *Trimmomatic* paired output file. The choice of the *Salmon* software lies in its ability to map reads directly to the transcriptome by ultra-fast read mapping procedure termed as quasi-mapping and to perform the quantification as well. It is a lightweight method suitable for RNA-seq reads, and it has the ability to correct for fragment GC-

content bias (Patro *et al.*, 2017). Moreover, the output files can be imported directly into the *DESeq2* software (Love, Huber and Anders, 2014) for differential expression analysis through the R Bioconductor package *tximport* (Soneson, Love and Robinson, 2016). This eliminates a few steps in the analysis, such as file type conversion and the use of separate quantification software. Moreover, *tximport* allows for both isoform (transcript) or gene-level expression analysis (Soneson, Love and Robinson, 2016).

2.3.2 Differential gene expression analysis of mapped and quantified reads

The quantified reads were imported via the *tximport* R package, and the differential expression analysis was performed using the R Bioconductor package *DESeq2* (Love, Huber and Anders, 2014; Soneson, Love and Robinson, 2016). Both *S. purpuratus* and *P. lividus* genome/transcriptome assemblies, in many cases, contain multiple transcripts belonging to the same gene. To be able to analyze data at the gene, and not at the transcript isoform level, the author of this thesis had generated custom annotation tables which were imported in the R working environment as "txt2gene" option of the tximport() function in the *tximport* R package (Soneson, Love and Robinson, 2016). *DESeq2* R package is equipped with methods to test for differential expression by using negative binomial generalized linear models. The estimates of dispersion and logarithmic fold changes incorporate data-driven prior distributions. *DESeq2* package is a method for differential analysis of count data, which uses the shrinkage estimation for dispersions and fold changes. This approach is based more on the strength than on the actual presence of differential expression (Love, Huber and Anders, 2014). Before running the differential analysis, batch effects were removed by including the batch factor and a condition (treatment) factor into the *dds* object. Differential analysis was performed running the *DESeq*() function. The R Bioconductor package *IHW* was used to filter the results of *DESeq*() output by including it as a filter in the results() function. Because the animals used in this study are collected in the wild from different populations, we cannot exclude their individual genomic makeup. Therefore, before performing further downstream

analysis, the samples were checked for potential variation using the Principal Component Analysis (PCA). The resulting output was rlog() transformed, and the PCAs were visualized using the ggplot() function in *ggplot2* package (Ginestet, 2011). Based on the clustering of different samples seen on the PCA plots, each sample belonging to the same cluster was given a batch name. IHW stands for "independent hypothesis weighting," which is a method that increases the power of large-scale multiple testing while controlling the false discovery rate (FDR), and it is a recommended approach for large data sets, such as high-throughput genomic data sets (Ignatiadis *et al.*, 2016). The results generated through the IHW function were then filtered by p-adjusted value ≤ 0.05. The P-adjusted value represents the p-value, which is transformed via Bonferroni correction. Bonferroni correction is the multiple-comparison correction used when multiple tests are performed at the same time. It is performed by dividing the significance levels by the number of comparisons (Bland and Altman, 1995). *S. purpuratus* p-adjusted value filtered datasets were annotated to SPU unique IDs and gene names using the WHL and SPU ID mapping table available from www.echinobase.org. The annotation Build8 was used to correctly annotate and subset transcription factors (TFs) and signaling molecules from the raw data sets. This annotation is available for download at www.echinobase.org, which is also utilized to build Echinobase whole database (Kudtarkar and Cameron, 2017). It contains functional annotation for 10542 genes and Gene Ontology terms annotation for 13672 genes. The author of this thesis customized these annotation tables by combining all GO terms and functional annotation with a WHL id number.

P. lividus p-adjusted value filtered datasets were annotated using the custom annotation made by Dr. Danila Voronov from Arnone's lab. The custom annotation was made by blasting the *P. lividus* proteome sequence to the *S. purpuratus* proteome sequence to their corresponding unique SPU IDs and *S. purpuratus* gene names (Tu, Cameron and Davidson, 2014). The final, filtered RNA-seq datasets were made by filtering out genes that were affected by standard MASO injection.

2.3.3 Gene ontology analysis of differentially expressed genes

The differentially expressed lists of gene SPU ids were uploaded and analyzed for Gene Ontology term enrichment with an online tool PANTHER (http://www.pantherdb.org/), that belongs to The Gene Ontology Consortium (http://www.geneontology.org).

2.4 DNA library preparation for ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing)

The DNA libraries were produced as described in Magri *et al*. Embryos of *S. purpuratus* grown at 15°C were collected at different developmental stages: 24hpf late mesenchyme blastula and 48hpf late gastrula. Two biological replicates of each stage were collected. An appropriate number of embryos was collected to obtain 100000 - 135000 nuclei: 540 embryos for the early gastrula stage and 270 embryos for the late gastrula stage. The embryos were centrifuged at 4ºC for 5 min at 500 rcf. They were washed twice with ice-cold artificial seawater (ASW). ASW was prepared by diluting 28.3 g NaCl, 0.77 g KCl, 5.41 g MgCl26H2O, 3.42 g MgSO4, 0.2 g NaHCO3, 1.56 g CaCl2 x 2H2O in 1L of dH2O and adjusted pH to 8.2. Each step was followed by 5 min centrifuge at 500 rcf and 4ºC. When the supernatant was removed, the embryos were lysed in 50 µl of lysis buffer using a pipette by pipetting up and down for 3-5 minutes, checking the lysis progress. When no debris is visible in the Eppendorf tube by the naked eye, the embryos are lysed. (Lysis buffer: 10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl2, 0.1% NP40). 25 µl of the lysed sample was taken and centrifuged for 10 minutes at 4°C at 500 rcf. The remaining 25µl of suspension was used for counting the nuclei. 1µl of DAPI (1:100) was added to this sample. The sample containing the dye was loaded into the Neubauer chamber, and nuclei were counted in the 25 µl under the microscope. While the sample was spinning, Transposition Reaction was prepared: 25 µl of 2x TD buffer (20 mM Tris(hydroxymethyl) aminomethane (Tris); 10 mM MgCl2; 20% (vol/vol) dimethylformamide, before the addition of dimethylformamide, pH was adjusted to

7.6 with 100% acetic acid), 23.75 µl nuclease-free water, 1.25 µl Tn5 (custom made enzyme provided by the Prof. J. L. Gómez-Skarmeta's lab).

After the centrifugation, the supernatant was removed as quickly as possible to avoid over-lysis, leaving only the pellet. The tagmentation reaction mixture was added, and the samples were incubated for 30 minutes at 37°C. Immediately following transposition, the sample was purified using a Qiagen MinElute Kit following the manufacturer's instructions. (5 µl of Sodium Acetate 3M was added before adding Buffer PB). The transposed DNA was eluted in 10 µl of Elution Buffer from the Qiagen MinElute kit (the minimal volume possible for the kit).

PCR Amplification of transposed DNA fragments. To amplify the transposed DNA fragments, PCR was performed. One PCR tube contained 10 µl of Transposed DNA, 10 µl of Nuclease Free H2O, 2.5 µl of Nextera PCR Primer 1* (10µM), 2.5 µl of Nextera PCR Primer 2* [Barcode] (10µM), 25 µl of NEBNext High-Fidelity 2x PCR Master Mix (New England Labs Cat #M0541). PCR cycles were set as follows:

1. 72 °C, 5 min
2. 98 °C, 30 sec
3. 98 °C, 10 sec
4. 63 °C, 30 sec
5. 72 °C, 1 min
6. Repeat steps 3-5, 15 times (the number was determined by qPCR)
7. Hold at 4°C.

Library quality check. The DNA concentration was measured using the Qubit dsDNA BR Assay Kit (Molecular Probes) instrument. Then 2-5 µl of the amplified DNA library (depending on concentration) was run on 2% Agarose gel. The presence of two bands at ~200 and ~400 base-pairs indicated the presence of one and two nucleosome DNA.

This experiment was conducted in collaboration with Marta Magri, a PhD student from Prof. Jose-Luis Gómez-Skarmeta group in Seville, Dr. Claudia Cuomo, a former Post-Doc, and Dr. Danila Voronov, from Dr. Arnone's laboratory. M.Magri, C. Cuomo,

D. Voronov, and the author of this thesis equally contributed to the experimental work.

2.5 ATAC-seq bioinformatics pipeline

The generated libraries were sequenced with the average sequencing depth of 57 million reads. The reads were trimmed with the *Trimmomatic* (Bolger, Lohse and Usadel, 2014) and mapped to the *S. purpuratus* Genome sequence V3.1 with *bowtie2* (Langmead and Steven L Salzberg, 2013), and the peaks were called with MACS2 software (Zhang *et al.*, 2008). This part of the analysis was conducted by Marta Magri, a PhD student from Prof. Jose-Luis Gómez-Skarmeta group in Seville, and Dr. Danila Voronov, from Dr. Arnone's laboratory (Magri *et al.*, in press). ATAC-seq data was intersected with ChIP-seq data to remove the false positive peaks and then analyzed as described in the 2.7 ChIP-seq bioinformatics pipeline by the author of this thesis.

2.6 ChIP-seq bioinformatics pipeline

Dr. Carmen Andrikou performed the *S. purpuratus* Brachyury Chromatin Immunoprecipitation experiments. The prepared libraries had two biological replicates for both 24hpf and 48hpf. The experiment was performed with two antibody preparations termed as #1 (for replicate #1) and #2 (for replicate #2). Dr. Danila Voronov performed the sequenced reads quality check, read mapping to the *S. purpuratus* Genome sequence V3.1, peak calling, replicate combination, and generated narrowPeak files. The author of this thesis performed downstream analysis starting from the NarrowPeak files.

ChIP-seq reads was mapped with *BWA mapper* (Li and Durbin, 2010) to v3.1 Genome sequence of *S. purpuratus*. The output was converted to bam format using *Samtools* (Li *et al.*, 2009), with the quality cut-off value of 30. The bam files were

converted to bed using *bedtools* bamtobed (Quinlan and Hall, 2010). The peaks were called with *Peakzilla* (Bardet *et al.*, 2013) with settings enrichment cut-off value of 0.5, FDR cut-off value of 1000, and Gaussian distribution option for model estimation. Cut-off value of 1000 was used to omit filtering of the generated peak files by the FDR value. The replicates were combined using the R package GenomicRanges (Lawrence et al., 2013) with fold enrichment score of 0.5 and fold enrichment of 2 in order to combine significant peaks by overlap. The fold enrichment score of 0.5 produced a much larger, less stringent dataset, while the enrichment score of 2 produced a smaller data set with the peaks of higher stringency.

All the subsequent analyses were conducted by the author of this thesis.

ChIP-seq and ATAC-seq intersection was accomplished using *bedtools* software with the default settings, including the *-wa* flag to keep the full ChIP-seq peak length and using the threshold of minimum 1 nucleotide overlap (Quinlan and Hall, 2010). The high stringency 24hpf ChIP-seq dataset was converted to FASTA format using the *getfasta* function in *bedtools* (Quinlan and Hall, 2010), and this file was used as an input file for *de novo* motif discovery. *De novo* motif discovery was conducted with default settings using *DREME*, part of the *MEME* suite (Bailey, 2011). *DREME* algorithm looks for enriched motifs in the datasets compared with shuffled sequences (Bailey, 2011). A newly discovered motif was searched for similarity to known motifs with *STAMP* online tool (Mahony and Benos, 2007; Mahony, Auron and Benos, 2007). Intersected ChIP-ATAC peaks were annotated and combined with RNA-seq data using *Homer* software (Heinz *et al.*, 2010). The *S. purpuratus* genome, transcriptome and ChIP-seq, ATAC-seq, and intersected ChIP-ATAC peaks were visualized on the *IGV* browser for quality check (Robinson *et al.*, 2011).

2.7 Protein sequence analysis

2.7.1 Sequence alignments

The multiple protein sequences were aligned using the *MUSCLE* algorithm implemented in *MEGA-X* software v10.1.8 (Edgar, 2004; Sudhir Kumar *et al.*, 2018). Pairwise sequence alignment was conducted with *Emboss Needle* (www.ebi.ac.uk/Tools/psa/emboss_needle/). The alignments were visually represented in *Jalview* v2.11.1.0 (Waterhouse *et al.*, 2009).

2.7.2 Prediction of protein secondary structure

The secondary structure predictions were conducted utilizing the online tool *RaptorX Property* (www.raptorx2.uchicago.edu/StructurePropertyPred/predict/) using template sequences discovered by the software itself. FASTA protein sequences were uploaded to the server, and the analysis was conducted automatically. The predicted structure was generated by the server using the multiple alignments method.

2.7.3 Prediction of ubiquitination, SUMOylation, and SUMO interaction sites

Prediction of ubiquitination sites was performed in a combined method using *UbPred* and *BDM-Pub* approach. Sequences in FASTA format were uploaded into the *UbPred* web-based server and *BDM-Pub* web-based server for the analysis (Li A, Gao X, Ren J, Jin C, 2009; Radivojac *et al.*, 2010)

UbPred is based on PSSM or Position-Specific Scoring Matrix, created by *PSI-BLAST* (Position-Specific Iterative Basic Local Alignment Search Tool) (Radivojac *et al.*, 2010). *PSI-BLAST* is used for the detection of distant relationships between proteins, and it is different compared to PAM and BLOSUM matrices, which give the same score to a substitution no matter where it appears (Bhagwat and Aravind,

2007). *UbPred* was developed using a set of new and known ubiquitination sites, and it has a calculated accuracy of 72%. Depending on the given score after the prediction, K residue is considered to be ubiquitinated if its score falls in the range between 0.62 and 1.00 (Visel *et al.*, 2009; Radivojac *et al.*, 2010).

BDM-PUB stands for Prediction of Ubiquitination site with the Bayesian Discriminant Method. It runs against a manually curated 260 experimentally validated ubiquitination sites of 154 different proteins, and it is based on a previously developed PK-specific phosphorylation site predictor (Xue *et al.*, 2006; Li *et al.*, 2009). The analysis was run with balanced cut-off settings.

SUMOylation and SUMO interaction site prediction was run locally by uploading FASTA formatted sequences into *GPS-sumo* software, and the analysis was run with low stringency settings (Zhao *et al.*, 2014). GPS or Group-based Prediction System algorithm runs with a database of 983 known SUMOylation sites in 545 proteins and 137 known SUMO interaction sites in 80 proteins (Ren *et al.*, 2009; Zhao *et al.*, 2014). All sequences used for proteomic analysis are listed in Table 2.1.

Table 2.1 The accession numbers and retrieval databases of used protein sequences.

Protein	Species	Accession Number	Database
Brachyury	*S. purpuratus*	SPU_013015.3a	www.echinobase.org
Brachyury	*H. pulcherrimus*	BAD74048.1	www.ncbi.nlm.nih.gov
Brachyury	*P. lividus*	CAD11971.1	www.ncbi.nlm.nih.gov
Brachyury	*L. variegatus*	AAL27986.1	www.ncbi.nlm.nih.gov
Brachyury	*S. mirabilis*	BAD74051.1	www.ncbi.nlm.nih.gov
Brachyury	*A. manni*	BAD74052	www.ncbi.nlm.nih.gov
Brachyury	*C. japonicus*	BAD74049.1	www.ncbi.nlm.nih.gov
Brachyury	*P. japonica*	BAD74050.1	www.ncbi.nlm.nih.gov
Brachyury	*P. pectinifera*	BAA84938.1	www.ncbi.nlm.nih.gov
Brachyury	*P. miniata*	PMI_009363.1	www.ncbi.nlm.nih.gov
Brachyury	*X. laevis*	AAH72031.1	www.ncbi.nlm.nih.gov
Brachyury	*M. musculus*	AAI20808.1	www.ncbi.nlm.nih.gov
Brachyury	*D. rerio*	XP_001343633.3	www.ncbi.nlm.nih.gov
Brachyury	*C. intestinalis*	AAD21079.1	www.ncbi.nlm.nih.gov
Brachyenteron	*D. melanogaster*	NP_524031.2	www.ncbi.nlm.nih.gov
Eomesodermin	*X. laevis*	NP_001081810.1	www.ncbi.nlm.nih.gov
Vegetally localized TF	*X. laevis*	AAB93301.1	www.ncbi.nlm.nih.gov
Smad1/5/8	*S. purpuratus*	SPU_023107.3a	www.echinobase.org
Smad1/5/8	*P. lividus*	Pliv02313.1	www.echinobase.org
Smad1/5/8	*P. lividus*	Pliv02431.1	www.echinobase.org
Smad1/5/8	*L. variegatus*	LVA_013523.1A	www.echinobase.org
Smad1/5/8	*L. variegatus*	LVA_013523.1B	www.echinobase.org
Smad1/5/8	*L. variegatus*	LVA_013523.1A_1	www.echinobase.org
Smad1/5/8	*L. variegatus*	LVA_013523.1B_2	www.echinobase.org
Smad1/5/8	*P. miniata*	PMI_019104.1	www.echinobase.org
Smad1/5/8	*P. parvimensis*	PPA_005373.1A	www.echinobase.org

2.8 Building a Gene Regulatory Network with Biotapestry

To represent complex interactions between transcription factors and signaling molecules at different developmental stages, the open-source, freely available graphical interface software *BioTapestry* was used. This program has been designed to aid in creating, updating, and sharing GRN models. Using *BioTapestry*, generated GRNs were organized temporarily and spatially (Longabaugh, 2012).

2.9 Whole-mount fluorescent immunohistochemistry

2.9.1 Protocol 1

The embryos were fixed in 4% PFA diluted in PEM buffer (100 mM PIPES, 5mM EGTA, 2mM MgCl2, 0.2% Triton X-100, pH to 6.8) for 5 min. Then, after the fixative was removed, they were washed once with 1X PBSTriton (0.2% Triton-X diluted in 1X PBS). After removing 1X PBSTriton, the embryos were washed a couple of times with 1X PBS and blocked in blocking solution (4% sheep serum, 1mg/ml BSA in PBST) for one hour at room temperature. Incubation with the custom-made primary antibody directed against Brachyury protein (anti-SpBra produced by Primm or anti-PlBra produced by GenScript) was in blocking solution overnight at 4°C. Anti-PlBra was used diluted 1:1500, and anti-SpBra 1:100 in blocking solution. The primary antibody was removed on the following day, and embryos were washed 5-6 times with 1X PBS. The secondary Alexa-Fluor™ 555 or 488 (ThermoFisher) antibody was diluted 1:1000 in blocking solution, and the embryos were incubated for 1 hour, followed by 4-5 times of washing with 1X PBS. Before mounting for observation under the microscope, DAPI was added diluted 1:10000.

2.9.2 Protocol 2

The embryos were fixed in 4% PFA prepared in FSW for 15 min. The fixative was removed, and the embryos were washed once with 1X PBST (0.1% Tween in 1X

PBS). After removing 1X PBST, the embryos were incubated for 1 min in ice-cold methanol. Methanol was removed, and the embryos were washed 4-5 times with 1X PBST. The blocking was performed over-night (or up to 3 days) at 4°C in a blocking solution containing 4% sheep serum and 1% bovine serum albumin (BSA). The custom-made primary antibody (anti-PlBra produced by GenScript) was diluted in blocking solution 1:250 and was added and incubated at 37°C for 1 hour and 30 min. The antibody was removed, and the embryos were washed 4-5 times in 1X PBST. The secondary Alexa-Fluor™ 555 or 488 (ThermoFisher) antibody was diluted 1:1000 in the blocking solution, and the sample was incubated for 1 hour at room temperature. After washing 4-5 times with PBST, DAPI was added diluted 1:10000, and the sample was mounted for imaging.

2.10 Imaging

For the live embryo imaging, the embryos were mounted in FSW on a microscope slide. They were observed under the Zeiss Axio Imager M1 microscope equipped with an Axiocam digital camera using the DIC mode. The images were taken by the Axiocam digital camera.

Embryos stained by whole-mount immunohistochemistry or whole-mount fluorescent hybridization were mounted and imaged under the Zeiss confocal laser scanning microscope LSM 700.

All images were processed using the ImageJ v1.49k software (Schneider, Rasband and Eliceiri, 2012). The analysis of the images was performed by Periklis Paganos and the author of this thesis.

2.11 Contribution statement

Periklis Paganos performed gamete collection, preparation of microinjection material, culture maintenance, RNA extraction, *in situ* hybridization, whole-mount immunofluorescence (Protocol 2) experiments, and imaging.

Dr. M. I. Arnone performed the microinjections.

Marta Magri, Dr. Claudia Cuomo, Dr. Danila Voronov, and the author of this thesis equally contributed to embryo culture maintenance and DNA library preparation for ATAC-seq experiments.

Marta Magri and Dr. Danila Voronov performed bioinformatics analysis of the ATAC-seq sequenced reads up to the peak calling.

Dr. Carmen Andrikou performed the ChIP-seq experiment.

Dr. Danila Voronov performed bioinformatics analysis of the sequenced ChIP-seq sequenced reads up to the peak calling.

The author of this thesis performed the rest of the work presented in this thesis.

CHAPTER 3

EXPLORING THE BRACHYURY PROTEIN STRUCTURE

This chapter contains the information resulting from studying the Brachyury protein structure based on the multiple sea urchin Brachyury protein sequence comparison and secondary structure predictions, ubiquitination site predictions, and predictions of protein-protein interactions. Possible causes of the peculiarity of the *S. purpuratus* protein temporal expression pattern are discussed throughout the chapter.

3.1 Introduction

Drafting of the Gene regulatory networks is based on representing the interactions between the nodes, which can be either transcription factors, signaling molecules, receptors, or terminal differentiation genes expressed in the same spatial domains at a specific time point. The common and most used approach is the spatiotemporal measuring of a specific node's mRNA levels considered to be included in a specific GRN. Since transcription factors are proteins themselves, and sometimes they do not share the same temporal expression pattern as their mRNAs, here, I propose that they should be included in drafting a GRN even if they do not share the same spatiotemporal domain as their corresponding mRNAs. This idea comes from the distinctive observation of the transcription factor Brachyury's temporal expression pattern in the sea urchin *S. purpuratus*.

3.1.1 S. purpuratus *Brachyury protein localization*

As already described in Chapter 1, the expression pattern of Brachyury in the sea urchin was found to be consistent in all species tested so far – *in situ* hybridization of *brachyury* in *S. purpuratus* (Rast *et al.*, 2002) and *P. lividus* (Croce, Lhomond and

Gache, 2001), and protein immunolocalization in *Lytechinus variegatus* (Gross and McClay, 2001) and *Arbacia lixula* (Andrikou, 2012).

However, a study in Arnone's group showed an unusual pattern of the Brachyury protein endodermal localization in *S. purpuratus* (SpBra) (Andrikou, 2012). Even if the *SpBra* mRNA is not detected anymore in the veg2 cells, after they ingress and form the archenteron, and are only detected in the veg1 endodermal cell lineage, immunofluorescence experiments revealed that the protein is retained in the veg2 cells, although in a lower amount than in the actively transcribing domains of oral ectoderm and blastopore region (Andrikou, 2012). The expression patterns of *brachyury* mRNA and the Brachyury protein localization in two sea urchin species, *S. purpuratus,* and *P. lividus,* are summarized schematically in Figure 3.1.

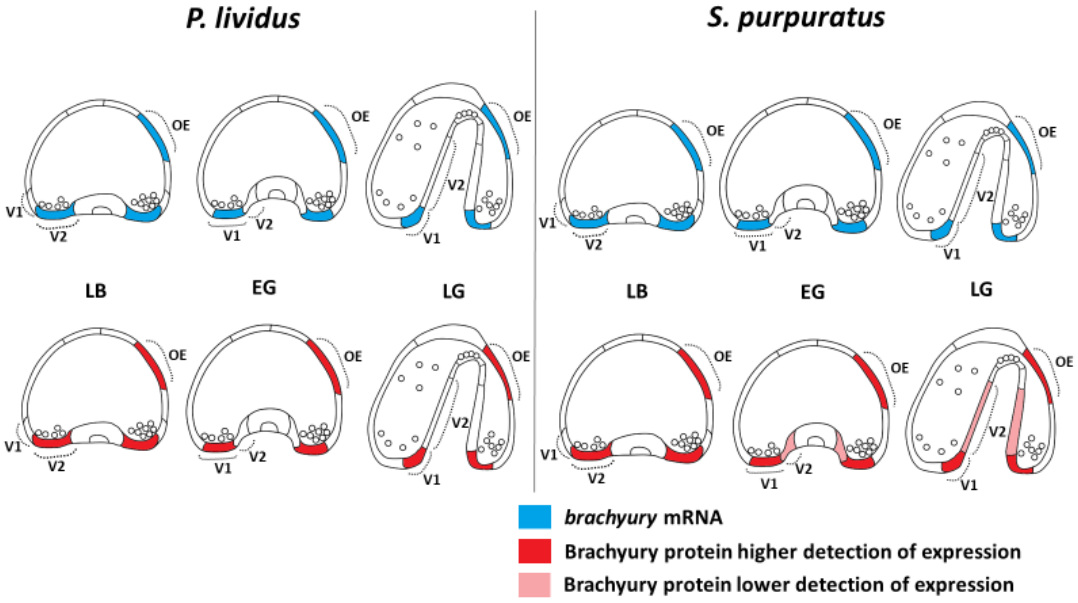


Figure 3.1 Comparison between mRNA and protein Brachyury localization in *P. lividus* and *S. purpuratus*, in the area of ectoderm and endoderm during three developmental stages. LB – late blastula, EG – early gastrula, LG – late gastrula, OE – oral ectoderm, V2 – veg2 endodermal linage, V1 – veg1 endodermal lineage. Note that SpBra is expressed in veg1 cells and in the veg2 cells after their invagination to form the archenteron, while PlBra is expressed only in the veg1 cells surrounding the blastopore. The localization of Brachyury in the oral ectoderm is shared between the two species.

This difference observed in *S. purpuratus* could be a novelty since there is a temporal overlap between mRNA and protein expression in all other species tested. Because SpBra is retained longer in the gut nuclei of *S. purpuratus*, it could potentially have a functional role. It could be involved in regulating specific genes that control the gut patterning, which could be different in *P. lividus*. The possible explanation for the endodermal presence of SpBra protein is that this protein could be more stable, thus retained in the gut, and involved in regulating the genes expressed in the veg2 derived archenteron cells of *S. purpuratus*. Another explanation could be that the *SpBra* mRNA is present in the archenteron cells of the 48hpf *S. purpuratus*, although in meager amounts and, therefore, undetectable by *in situ* hybridization. The following subsections will concern the first explanation and the factors that could contribute to the differential stability of *S*. *purpuratus* Brachyury protein.

3.1.2 Protein structure and its effect on stability

Each protein is built from different combinations of the 20 amino acids connected with peptide bonds among each other, and by the four levels of the structural organization – the primary or amino acid sequence, the secondary or the folding of a peptide chain into alpha-helices or beta-sheets, the tertiary or three-dimensional structure of interacting amino acids in a polypeptide chain, and quaternary structure or macromolecular structure formed by different interactions between the polypeptide chains.

Amino-acids, the building blocks of proteins, are classified into different groups based on structural, chemical, and physical properties, and traditionally, the classification is based on Taylor's model (Taylor, 1986). Taylor has grouped all amino acids based on two categories: size and water affinity. Based on the size, all amino acids are grouped in large, small, or tiny category, based on water affinity into hydrophilic (which can be neutral, positively or negatively charged) and hydrophobic (which can be aromatic if they contained an aromatic ring, or aliphatic). After the polypeptide chain is synthesized, it goes through a series of changes that affect its folding, bringing it to a native, functional state. Protein stability is defined as a net

balance between free energies in the folded and unfolded states. The folded state is stabilized through disulfide bonds and non-covalent interactions (hydrophobic, electrostatic, hydrogen bonding, and van der Waals forces), and the unfolded state is influenced by entropy (Gromiha, Nagarajan and Selvaraj, 2019). One of the most important mechanisms that shape and bring the protein to its native state and make it stable is the post-translational modifications (e.g., phosphorylation, glycosylation, hydroxylation, acetylation, ubiquitination, etc.), which are known to be highly specific towards amino acids (Feher, 2012; Krishna & Wold, 2006; Parekh & Rohlff, 1997).
The biological function of the protein depends on its chemical and physical properties. These properties are caused by amino acid residue interaction in a three-dimensional space. Changes in amino acid residues can have a strong impact on biochemical mechanisms and evolution (Starr and Thornton, 2016). Therefore, if a specific amino acid is substituted for another, or an extra residue is added to or removed from a polypeptide chain, the overall protein structure can be changed. This means that the "new protein" can also be subject to novel post-translational modifications. For example, serine, threonine, and tyrosine are often phosphorylated, while lysine is the amino acid known to be affected by ubiquitination (Krishna and Wold, 2006).

3.1.3 Post-Transcriptional Modifications: Ubiquitination and SUMOylation of the T-box proteins

T-box transcription factors are dynamically expressed, often prone to different post-translational modifications (PTMs) that can influence their binding preferences, stability, and activity. Some of the most important PTMs are ubiquitination and SUMOylation.

Ubiquitination (ubiquitylation) is an enzymatic PTM that includes a three-step mechanism: activation by E1 enzymes (ubiquitin-activating enzymes), conjugation by E2 enzymes (ubiquitin-conjugating enzymes), and transfer of the E3 enzyme (ubiquitin ligase) to its protein target (Passamore and Barford, 2004). Ubiquitin is a protein that binds its targets very selectively to the lysine residues (Passamore and Barford, 2004). Ubiquitination is a reversible reaction; however, it is one of the main

processes that lead to protein degradation or proteolysis (Passamore and Barford, 2004). Reversible ubiquitination is usually involved in different signaling pathways such as transcription, DNA repair, cell cycle, immune response, and different protein-protein interactions (Hofmann and Pickart, 1999, 2001; Deng *et al.*, 2000; Spence *et al.*, 2000; Wang *et al.*, 2001; Schnell and Hicke, 2003).

SUMOylation is post-translational modification analogous to ubiquitination. Small ubiquitin-like modifier or SUMO is a family of proteins that can attach to lysine residues through step-by-step enzymatic processes (Gareau and Lima, 2010). SUMOylation involves SUMOconjugation through the action of E1, E2, and E3 enzymes. SUMO precursors need to be cleaved via sentrin/SUMO-specific protease or SENP enzymes before being able to be conjugated. SUMOylation is a highly dynamic process, and deSUMOylation is performed through SENP enzymes as well (Gareau and Lima, 2010). SUMOylation pathway is vital for the protein localization and transport, regulation of gene expression, protein stability, genome maintenance, cell cycle, and stress response (Gill, 2004, 2005; Hay, 2005)

Previous studies have shown that canonical SUMOylation consensus motif is ψ–K–X–E, where ψ is any hydrophobic amino acid (A, I, L, M, P, F, V or W), and X is any amino acid residue (Rodriguez, Dargemont and Hay, 2001; Sampson, Wang and Matunis, 2001). On the contrary, Xue's group experimental data shows that only 40% of all SUMOylation sites do not follow the rule, and it is a field to be further explored (Zhao *et al.*, 2014).

There have been a couple of studies on ubiquitination and its involvement in T-Box protein stability, precisely of the T-bet transcription factor. T-bet is the key transcription factor involved in the differentiation of the T-helper (Th1) cells. It has been shown that lysine at position 313 (K-313) is crucial for T-bet protein stability. K-313 controls the T-bet stability through ubiquitination and proteasomal degradation. Moreover, this residue is directly involved in DNA binding to the IFN-γ gene promoter and phosphorylation of threonine-302 (T-302) that allows it to suppress NFAT1 activity (Jang *et al.*, 2013). Pan and colleagues showed some evidence that USP10 (Ubiquitin carboxyl-terminal hydrolase 10) could be

responsible for T-bet stabilization through deubiquitination that prevents proteasomal degradation and enhances IFN-γ secretion (Pan *et al.*, 2014).

T-box TF SUMOylation has been intensively studied in *C. elegans*. It has been shown that SUMOylation of the TBX-2 TF is required in *C. elegans* development of pharyngeal muscles. Roy Chowdhuri and colleagues have identified two consensus SUMOylation sites in TBX-2 protein: LKIE (230) and VKKE (399) that are SUMOconjugated via UBC-9 enzyme (Roy Chowdhuri *et al.*, 2006; Crum and Okkema, 2007). Moreover, there is some evidence that TBX-2 acts as a transcriptional repressor after SUMOylation (Crum and Okkema, 2007). The most recent TBX-2 study has shown that SUMOylation is tightly connected to TBX-2 ability to autoregulate its expression via a negative loop (Milton and Okkema, 2015). It seems that SUMOylation is a conserved mechanism that regulates T-box activity (Huber *et al.*, 2013). Human TF TBX-22, which is essential for craniofacial development, is modified by SUMOylation, which allows it to act as a repressor (Andreou *et al.*, 2007).

A study on the involvement of TBX3 TF in melanoma progression has shown that AKT3 (AKT Serine/Threonine Kinase 3) directly acts on TBX3 by phosphorylating serine at position 720, having a massive effect on its stabilization, nuclear localization, and repression of E-cadherin (Peres, Mowla and Prince, 2015).

3.1.4 Interaction between Smad and T-box proteins

Smad transcription factors are an essential part of the transforming growth factor-β (TGF-β) signaling pathway during metazoan development (Shi and Massagué, 2003). TGF-β family includes TGF-β, nodals, activins, bone morphogenic proteins (BMPs), and others (Massagué, Seoane and Wotton, 2005). In vertebrates, Smad TFs are classified into R-Smads, co-Smads, and I-Smads. R-Smads or receptor-regulated Smads are substrates for BMP receptors, and, in vertebrates, they are directly phosphorylated and include Smad1, Smad2, Smad5, and Smad8. Smad4, also known as the Co-Smad, is a functional partner of all other Smads. I-Smads or inhibitory Smads involve Smad6 and Smad7; they negatively regulate TGF-β signaling by competing with Co-Smad or R-Smads (Massagué, Seoane and Wotton,

2005). Smad proteins are formed out of three structural elements: the conserved N-terminal MH1 domain (absent in I-Smads), the C-terminal MH2 domain, and the structurally variable linker domain (Shi and Massagué, 2003).

Different T-box TFs have different preferences for inducing specific genes. For instance, in vertebrates, three different mesodermal T-box factors seem to be very picky: VegT (Antipodean) and Eomesodermin (Eomes) can induce dorso-anterior marker *goosecoid*, while Brachyury (Bra) cannot. Moreover, Bra can induce both *Wnt11* and *Bix4*, while VegT and Eomes cannot (Conlon *et al.*, 2001). For a long time, the cis-regulatory elements binding preferences of T-Box TFs were thought to be regulated by the slight differences in T-box domain structure (Conlon *et al.*, 2001), but the T-box domain is highly conserved; and now we know that the differences in binding preferences are most likely determined by the variable structure in T-box TFs' N-terminus and C-terminus domain (Messenger *et al.*, 2005). It has been shown that the inducing capacity of Bra is mediated by its ability to interact with the C-terminal (MH2) domain of Smad1 TF through an N-terminal HLL(S/N)AV sequence adjacent to the T-box domain of Bra (Marcellini *et al.*, 2003; Messenger *et al.*, 2005; Marcellini, 2006). It seems that the interaction is particularly strong between Brachyury's H-L-L, A-V, and Smad1's S-378, Y-336, V-338, and T-343, respectively (Marcellini, 2006). As previously described in the introductory chapter of this thesis, Bra function is context-dependent (Faial *et al.*, 2015). In other words, the specific binding of Bra to its target genes is tightly regulated by the signaling it receives (Activin/Nodal or BMP4) and the Smad (Smad2/3 or Smad1) co-factor it operates with (Faial *et al.*, 2015). In addition, it has been shown that Bra and Smad1 are able to synergistically activate neural repressor *ventx1.1* through both BMP4/Smad1 and Bra/FGF signaling in *Xenopus* embryos (Shiv Kumar *et al.*, 2018). Inhibitory roles of Smads have been studied in T-box factors as well. Namely, there is some evidence that Smad6 can inhibit *Tbx6* transcriptional activity by recruiting Smurf1, a known ubiquitin E3 ligase (Chen *et al.*, 2009).

Exploring the composition of a protein primary structure can give much information about putative post-translational modifications, the possible physical interactions that lead to secondary structure (forming alpha-helices or beta-sheets), and, finally, allows to predict the three-dimensional structure. Moreover, the protein sequence

analysis could show potential protein-protein interaction sites based on the validated published data from phylogenetically related proteins. Therefore, this Chapter aims to perform different in silico protein sequence level analyses of different echinoderm Brachyury protein homologs in order to identify features that could lead to specific protein-DNA interactions and protein-protein (specifically interaction with Smad proteins) and to predict putative PTMs that could explain the proposed differences in protein stability of SpBra.

3.2 Results and discussion

3.2.1 Brachyury protein immunolocalization in S. purpuratus *and* P. lividus

As mentioned previously, the veg2 endoderm derived gut cells' show localization of Brachyury protein in the 48hpf late gastrula of *S. purpuratus* (SpBra). The immunofluorescence experiment was repeated in both *S. purpuratus* and *P. lividus* using the antibody designed against SpBra by Andrikou.

The experiment confirmed that the Bra protein's gut localization was only found in *S. purpuratus* (Figure 3.2). This antibody was designed using the sequence domain of the SpBra protein adjacent to the DNA-binding domain (Figure 3.3).

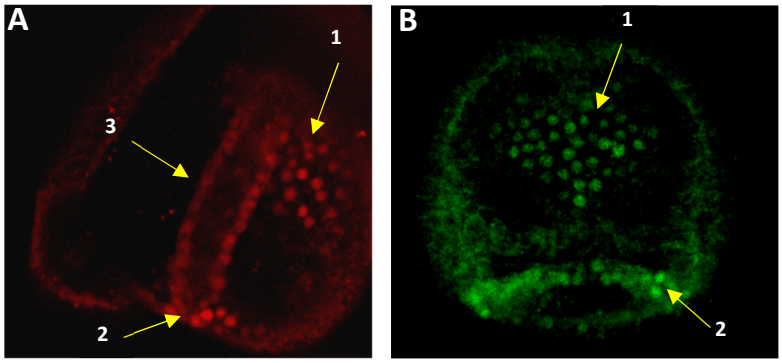


Figure 3.2 Brachyury protein immunolocalization in *S. purpuratus* and *P. lividus* at the late gastrula stage.

A) whole-mount immunohistochemistry (WMIHC) with anti-SpBra in *S. purpuratus,* lateral view; B) WMIHC with anti-SpBra in *P. lividus,* oral view. SpBra protein is localized in the oral ectoderm (1), the ring of cells around the blastopore (2) and the veg2 derived cells of the archenteron (3). PlBra protein is localized in the oral ectoderm (1) and in the ring of cells around the blastopore (3).

However, since the antibody used in previous experiments was generated against the SpBra peptide sequence, and it was not PlBra specific, another antibody was designed against the PlBra sequence (anti-PlBra). The design was somewhat different, and the selected protein sequence was longer, spanning from the N-terminus and including the DNA-binding domain of the PlBra protein (Figure 3.3). Since this antibody was generated against the most conserved domain (T-box DNA-binding domain), it was expected to recognize the protein in both species with similar specificity. This antibody was tested against the Bra proteins in both species using two different protocols described in Chapter 2. The results of the testing are shown in Figure 3.4A. The PlBra protein localization was corresponding to the *Brachyury* mRNA expression pattern – the oral ectoderm and the ring of cells surrounding the blastopore. Contrary to the observation made using the anti-SpBra, anti-PlBra could not localize the protein in the archenteron in *S. purpuratus* (Figure 3.4C). The immunohistochemistry experiments using the anti-PlBra antibody were repeated on *S. purpuratus* implementing Protocol 2 (see Chapter 2 for details), and it gave the same result as Protocol 1 – the absence of the protein staining in the gut (Figure 3.5).

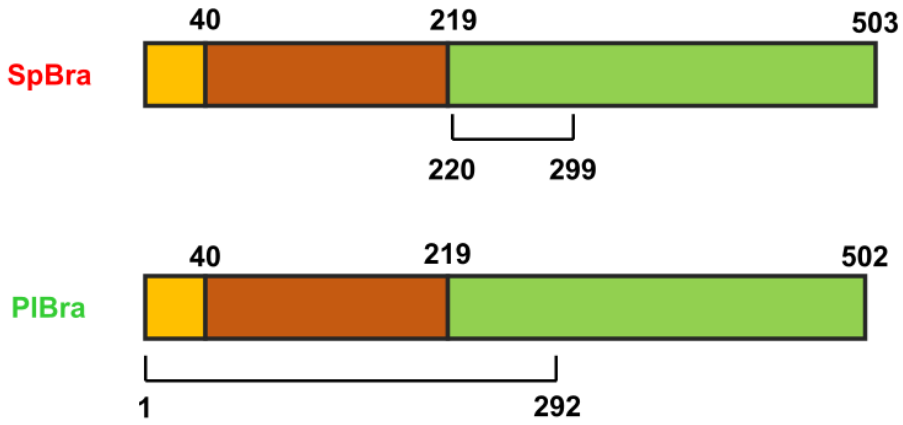


Figure 3.3 Schematic representation of the SpBra and PlBra protein sequence and the regions used to generate the anti-SpBra and anti-PlBra antibodies. The N-terminal domain is represented in yellow, the DNA-binding domain (T-box) in brown the C-terminal domain in green. The region used to generate the anti-SpBra was 79 aa long and adjacent to DNA-binding domain in the C-terminal domain (Andrikou, 2012), while the amino acid sequence that was used to generate anti-PlBra was 292 aa long and it contained N-terminal region, DNA-binding domain and 73 aa in the C-terminal region.

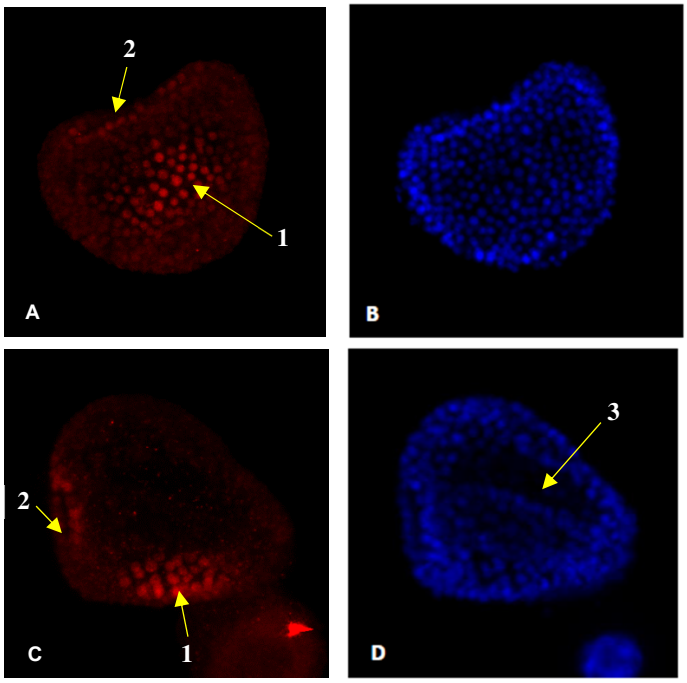


Figure 3.4 Immunolocalization of Brachyury protein in the nuclei of *P. lividus* and *S. purpuratus* detected with the anti-PlBra antibody using the Protocol 1. (A) Detection of PlBra in *P. lividus,* oral view (B) nuclear staining with DAPI in *P. lividus,* oral view (C) Detection of SpBra in *S. purpuratus,* lateral view (D) nuclear staining with DAPI in *S. purpuratus,* lateral view; 1 - oral ectoderm, 2 - blastopore, 3 – archenteron.

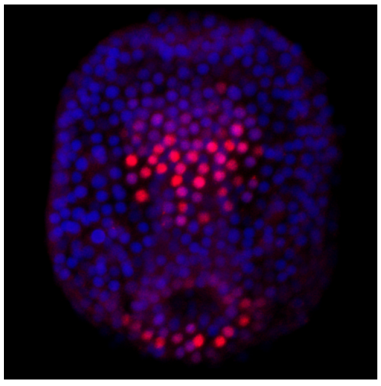


Figure 3.5 Immunolocalization of Brachyury protein in the nuclei of *S. purpuratus* detected with the anti-PlBra antibody using the Protocol 2. Brachyury staining is in red and nuclear DAPI staining is in blue. SpBra protein is localized in the oral ectoderm and a ring of cells surrounding the blastopore. Courtesy of Periklis Paganos.

This observation that no SpBra could be detected in archenteron of *S. purpuratus* using anti-PlBra antibody could be explained by the possibility that the anti-PlBra antibody has lower sensitivity to the purple sea urchin protein compared to the anti-SpBra, which is specific to the SpBra protein, and it is not able to detect protein in the gut. The hypothesis of anti-SpBra antibody high sensitivity can be supported by the observation that the gut protein presence using the anti-SpBra is not at the same level since the signal in the gut is rather dim, while the same anti-SpBra antibody gives a much brighter signal in the oral ectoderm and blastopore region signal (Figure 3.2A).

3.2.2 Brachyury protein sequence evolution: different mechanisms affecting protein stability, structure and interaction with other molecules

To discover putative important amino acid substitutions responsible for the difference in the Brachyury protein stability in *S. purpuratus* and *P. lividus*, the protein sequences were aligned. In addition, computational predictions of the protein secondary structure were conducted in order to detect if any regions of the SpBra protein could have preferences for the different protein folding or protein-protein interactions compared to the sequences of other sea urchin Brachyury proteins.

Protein sequence comparison between echinoderms

To discover putative important amino acid substitutions responsible for the difference in the stability of the transcription factor Brachyury of *S. purpuratus* compared to *P. lividus*, PlBra, and SpBra sequences were aligned with *Emboss Needle*. Brachyury's primary structure is highly conserved among *S. purpuratus* and *P. lividus* with 91.8% identity, 95.6% similarity, and 1 gap. The total number of different amino acid residues between *P. lividus* and *S. purpuratus* is 40. SpBra sequence contains an additional amino acid asparagine (N) in position 254. The pairwise alignment is present in the Non-book component of this thesis.

Other echinoderm species' Brachyury sequences were included in the analysis and aligned using the *MUSCLE* program in order to identify potentially important substitutions (Edgar, 2004). The reason behind this is that the expression of Brachyury follows a similar pattern in all the species tested so far. For example, the localization of the *L. variegatus* Brachyury (LvBra) protein was found to be identical to the localization of *P. lividus* (Gross and McClay, 2001; Andrikou, 2012). There is no protein expression data available for *H. pulcherrimus*, but the mRNA localization is in both oral ectoderm and the ring of cells surrounding the blastopore (Hibino *et al.*, 2004). Sand dollars *P. japonica*, *C. japonicus brachyury* mRNA expression was found in oral ectoderm and blastopore, while *A. manni* and *S. mirabilis* seem to have an additional third domain of expression, which is the oral side of the archenteron in the mid-gastrula stage of development (Hibino *et al.*, 2004). Starfish species *P. pectinifera* and *P. miniata* have two conserved domains of expression, the oral ectoderm and the ring of cells surrounding the blastopore (Shoguchi, Satoh and Maruyama, 1999). In addition, *H. pulcherrimus* is phylogenetically the closest species related to *S. purpuratus*, and the Bra protein localization likely follows the same pattern with the third domain of expression in the gut. For this reason, the residues that distinguish *S. purpuratus* and *H. pulcherrimus* from other echinoderm Brachyury proteins were considered relevant and will be discussed below. This 'larger' comparison narrowed down the search to 14 putatively important residues that might affect the protein structure and behavior of SpBra (Figure 3.6). The full alignment is available as a part of the Non-book component of this thesis. Therefore, the amino acid residues different from *P. lividus* but identical to *S. purpuratus* (except for *H. pulcherrimus*) were considered less likely to contribute to the different stability of the SpBra protein compared to PlBra protein.

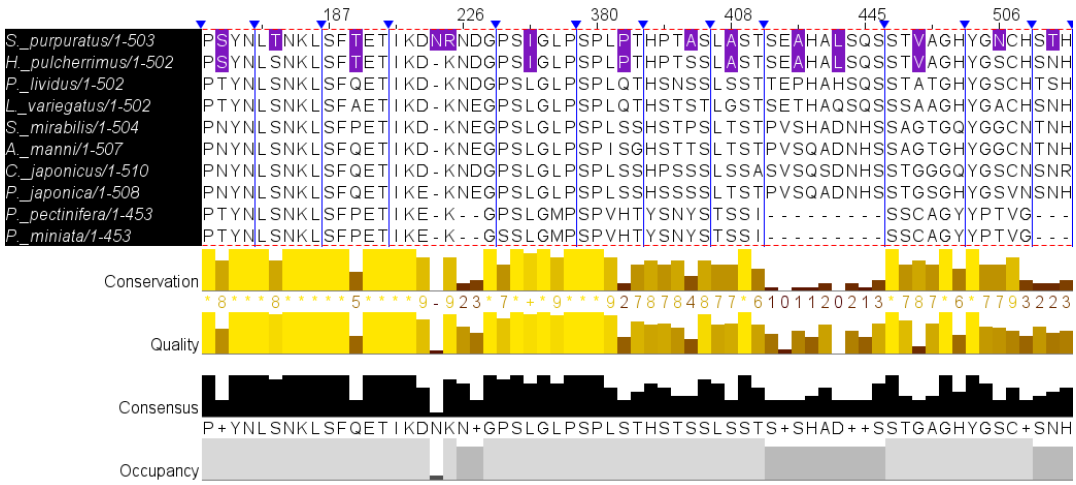


Figure 3.6 Non-conserved amino acid residues in *Strongylocentrotus purpuratus* and *Hemicentrotus pulcherrimus* Brachyury proteins that might be responsible for the differences in protein stability, protein-DNA and protein-protein interactions. The figure represents multiple sequence alignment of the sea urchin (*S. purpuratus*, *H. pulcherrimus, P. lividus, L. variegatus)*, sand dollar (*S. mirabilis, A. manni, C. japonicus, P. japonica*) and starfish (*P. pectinifera*, *P. miniata)* Brachyury protein sequences. The sequence length is written next to the species name. Some protein regions are hidden for a better visual representation, and are marked with blue arrows. Below the alignments, calculated level of conservation (-,0,1,2,3,4,5,6,7,8,9 and *), alignment quality (based on BLOSSUM62 scores), consensus sequence and occupancy of each amino acid are shown (1-10). Putative important amino acid substitutions of *S. purpuratus* are marked with purple color. One residue is located in the N-terminal region: S-13, two of them in the T-box DNA-binding domain: T-148 and T-189, and 11 residues are located in the C-terminal domain: N-224, R-225, I-268, P-369, A-380, A-387, A-413, L-416, V-431, N-472, and T-477.

The first observed substitution is the residue 13, where threonine (T) was exchanged with serine (S). Both serine and threonine are small, polar amino acids that often substitute with each other without having consequences on the protein structure, and they are both prone to phosphorylation (Taylor, 1986; Betts and Russell, 2007). However, the importance of serine is often neglected. It has been shown that serine has an essential role in metabolism (Kalhan and Hanson, 2012). Serine is the major source of one-carbon units for methylation reactions (Kalhan and Hanson, 2012). Moreover, if found in protein functional centers, the substitution with threonine can affect the stability of the protein (Smeda *et al.*, 1993; Masabni and Zandstra, 1999).

Residue 148 showed quite the opposite. In all sequences analyzed, this position is occupied by serine (S), while in *S. purpuratus*, it was substituted by threonine (T). Since this position is located in the DNA-binding domain, it might be possible that it could change SpBra's preference for the chromatin binding. Similarly, still in the T-box region, residue 189 seems to be conserved and represented by glutamine (Q), whilst in *S. purpuratus* and *H. pulcherrimus*, this residue is substituted by threonine (T). These two substitutions in the DNA binding domain could be affecting protein overall structure by the possible phosphorylation of threonine (Betts and Russell, 2007). All other potentially relevant substitutions for the differences in SpBra structure are found in the protein's C-terminal region. Closest to the T-box domain, SpBra protein has acquired an extra asparagine at position 224. Asparagine (N) is a polar amino acid often affected by the N-linked glycosylation (Gavel and Heijne, 1990; Krishna and Wold, 2006). The acquirement of multiple asparagine residues could be quite important for the proper folding of a protein, affecting its stability (Breitling and Aebi, 2013). Moreover, at position 225, there was a substitution from lysine (K) to arginine (R). This change from lysine (K) to arginine (R) could be responsible for resistance to ubiquitination and SUMOylation, that have been known to occur at lysine residues (Passamore and Barford, 2004; Gareau and Lima, 2010). Arginine might be post-translationally modified by methylation, which could affect the interaction with substrate proteins (Fulton, Brown and George Zheng, 2019). There is isoleucine (I) to leucine (L) substitution at position 268, which is most likely synonymous. These amino acids are both non-reactive hydrophobic and aliphatic, and mostly affected by hydroxylation and acetylation and the consequence of these mutations is, most likely, minimal towards the protein stability (Taylor, 1986; Krishna and Wold, 2006; Betts and Russell, 2007; Feher, 2012). However, these substitutions could have an important role in substrate recognition and binding of hydrophobic ligands (Betts and Russell, 2007). Moreover, this substitution might have evolutionary importance since it occurred in both *S. purpuratus* and *H. pulcherrimus*, the closest echinoderm relatives in the evolutionary tree compared to the other species analyzed. A similar observation was made for the residues at the positions P-369 (equivalent to P-368 in *H. pulcherrimus*), A-387 (A-386 in *H. pulcherrimus*), A-413 (A-412 in *H. pulcherrimus*), L-416 (L415 in *H. pulcherrimus*), and V-431 (V-430 in *H. pulcherrimus*). Proline (P) is an amino acid that has quite an

unusual geometry; it is a unique small amino acid whose side chain is connected to the protein backbone twice, forming a five-membered ring; it is significant for protein folding since it is found in tight turns, where the polypeptide chain change its direction (Betts and Russell, 2007). In addition, it has been shown that proline plays a role in stabilizing the proteins (Kotake *et al.*, 2005; Prajapati *et al.*, 2007). Studies have linked mutations that replaced serine with alanine to the changes in the protein-protein interactions and the increased biological activity (Stockhaus *et al.*, 1992; Coley *et al.*, 2000). Since serine to alanine substitution appeared three times in the same region of the SpBra protein (A-380, A-387, and A-413), it is highly probable that it has a stabilizing effect. At positions 413 (412) and 431 (430), both *S. purpuratus* and *H. pulcherrimus* have leucine (L) and valine (V), respectively. These amino acids have similar properties, and they could be both affected by hydroxylation or acetylation, which could have an effect on overall protein three-dimensional structure (Taylor, 1986; Krishna and Wold, 2006; Betts and Russell, 2007; Feher, 2012). Last but not least, at the very end of the C-terminal domain, two substitutions appeared in *S. purpuratus*, which are not seen at those locations in other echinoderm Bra proteins: asparagine (N-472), which replaced the consensus serine (S) and threonine (T-477) which replaced the consensus asparagine (N).

In conclusion, these results show that the *S. purpuratus* Brachyury sequence is indeed different and that there might be a mechanism involved in stabilizing the SpBra protein through some newly acquired or substituted amino acid resides. In particular, the acquiring of the additional asparagine residue could have an effect on the folding of the SpBra protein that can affect its overall stability. Moreover, the regions that could be interesting to explore are the ones that contain multiple consecutive substitutions, especially in the C-terminal domain.

Secondary structure predictions

To explore the structure of the protein sequences more deeply, considering the interactions between the adjacent amino acid residues, I looked at the predicted secondary structure models of the Brachyury proteins of *S. purpuratus*, *P. lividus*, and *L. variegatus*. These three species were selected for comparison due to their known protein localization.

To predict the secondary structures that polypeptide chains can form, the multiple-template threading method was used to align a single protein sequence to multiple templates (Peng and Xu, 2011). To this aim, the analyses were run on the web-based protein structure prediction software *RaptorX Property* (Peng and Xu, 2011; Feher, 2012; Källberg *et al.*, 2012; Ma *et al.*, 2012, 2013; Zhu *et al.*, 2018). The secondary structure of PlBra, LvBra, and SpBra are shown in Figures 3.7. High confidence prediction was observed in the region spreading from residue 11 to residue 212 (Figures 3.7). This is not surprising, considering that the conserved DNA binding domain is situated there. This prediction method is template-based, where the xBra template was found to be the closest match. xBra is the published crystallography structure coming from the Brachyury T-box binding domain of *Xenopus laevis* (Müller *et al.*, 1997). For PlBra, the model was created with a p-value of 5.11e-17; only 6% of the sequence was found to be an alpha-helix, 17% beta-sheet, and the majority was considered as a random coil (75%). LvBra prediction was created with a p-value of 3.05e-17; 6% of the sequence was predicted to form an alpha-helix, 18% beta-sheet, and 74% a random coil structure. A similar situation was observed with SpBra, where the p-value was 5.59e-17; with 6% alpha-helix, 18% beta-sheet, and 74% random coil predictions. The particular regions of interest are marked with green color, depicted in Figure 3.8.

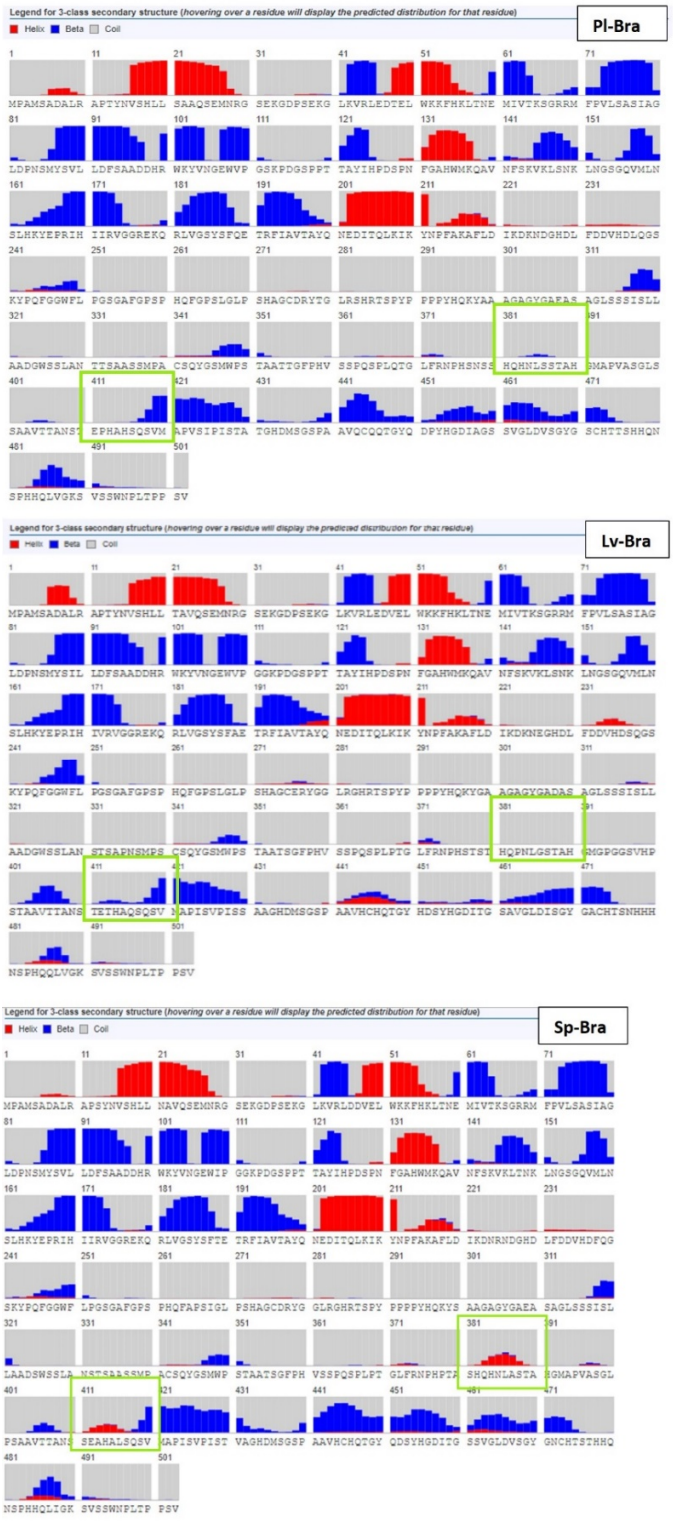


Figure 3.7 *Paracentrotus lividus*, *Lytechinus variegatus* and *Strongylocentrotus purpuratus* Brachyury protein secondary structure predictions. Colored bars represent chance of specific amino acid residue to form a helix (red), a beta-sheet (blue) or a random-coil (gray). The regions marked with the green rectangles show detectable differences between SpBra and the two other sea urchin species.

It seems that only SpBra protein tends to form alpha-helix structures in the two regions, specifically the region between the residues 381 and 391 and the region between the residues 411 and 421. These regions in *P. lividus* and *L. variegatus* Bra sequences do not show any particular preferences in forming neither alpha-helix nor beta-sheet, which is visible by the gray color representing the random coil structure, or the equal possibility to form any of the mentioned secondary structures (Figure 3.8).

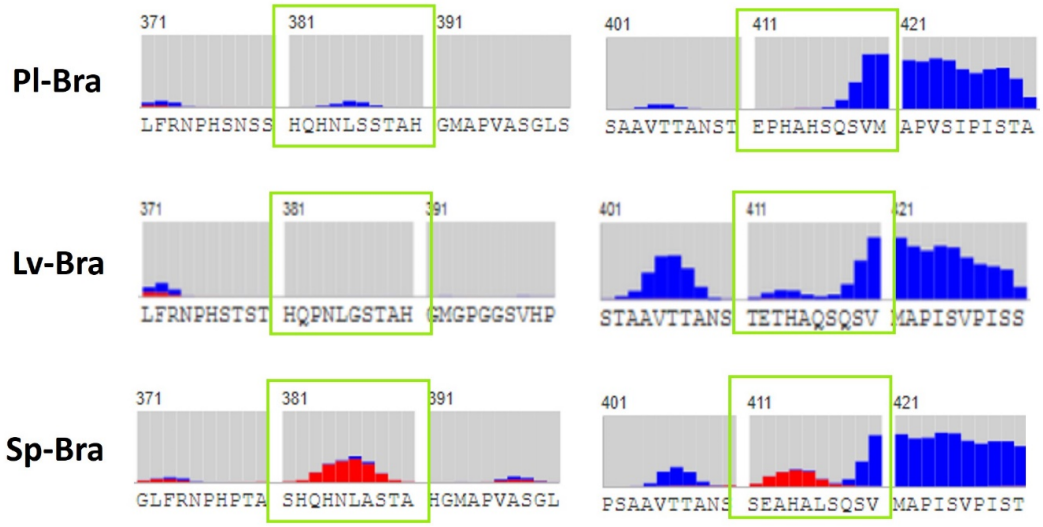


Figure 3.8 Comparison between the three different C-terminus domain secondary structure prediction results in *Paracentrotus lividus*, *Lytechinus variegatus*, and *Strongylocentrotus purpuratus* Brachyury proteins.

PlBra: *P. lividus* Brachyury, LvBra: *L. variegatus* Brachyury, SpBra: *S. purpuratus* Brachyury**.** Green rectangles mark sequence regions that show different folding predictions. Blue colored bars represent the probability of a residue forming a beta- sheet, red colored bars represent the probability of a residue forming an alpha-helix and gray areas represent an equal possibility of forming both.

This finding is interestingly connected to the localization of the multiple substitutions that are described in the paragraph above, where *S. purpuratus* had acquired proline (378), alanine (380), alanine (387), in the first alpha-helix enriched region; and serine (411), alanine (413), and leucine (416) in the second alpha-helix enriched region. It could be possible that this conformation preference is affected not just by those 6 residues but by their interaction with the adjacent residues as well, which

could all be subject to multiple post-translational modifications. These two regions might affect SpBra protein stability compared to the other species. Moreover, these regions are located in the C-terminal domain, and it has been shown that C-terminus is involved in target binding in some transcription factors and that the binding activity is influenced by the phosphorylation of specific residues (Singh *et al.*, 2018). In addition, phosphorylation is associated with determining if the transcription factor functions as an activator or a repressor, and this post-translational modification can affect its overall stability (Tootle and Rebay, 2005).

3.2.3 Predicting ubiquitination and SUMOylation sites in different sea urchin Brachyury proteins

To discover amino acid residues that could lead to higher protein stability of SpBra, compared to other echinoderm species, computational predictions of ubiquitination, SUMOylation, and SUMO-interaction sites were performed using the *UbPred*, *BDM-PUB*, and *GPS-sumo* open-source software (Li *et al.*, 2009; Ren *et al.*, 2009; Radivojac *et al.*, 2010; Zhao *et al.*, 2014).

Ubiquitination sites prediction was performed in four sea urchin species, S. purpuratus, *H. pulcherrimus*, *P. lividus*, and *L. variegatus*, with a combined approach.

As shown in Table 3.1, some lysine sites were predicted positive for ubiquitination with both software algorithms, while some sites did not pass the threshold. However, the combined method was used, and the result was reported positive if it appeared in at least one dataset. Based on the combined predictions of UbPred and BDM-PUB, *H. pulcherrimus*, *P. lividus* and *L. variegatus* Bra proteins were positive for 13 ubiquitination sites K-33, K-39, K-42, K-113 (equivalent to K-112 in *L. variegatus*), K-144 (K-143 in *L. variegatus*), K-146 (K145 *L. variegatus*), K-150 (K-149 in *L. variegatus*), K-210 (K209 in *L. variegatus*), K-222 (K221 in *L. variegatus*), K-241 (K-240 in *L. variegatus*), K-297 (K-296 in *L. variegatus*) and K-489. The highest scores were for K39 and K42, located in the N-terminal domain, and K-42, located just at the beginning of the T-box domain. All other residues situated in the T-box domain and C-terminus had a lower score and/or were detected as positive in only one

dataset. All tested species, apart from *S. purpuratus*, had a positive ubiquitination site at K-297. The interesting observation was that in *S. purpuratus*, Brachyury protein has one extra residue of asparagine (N-224). Moreover, the consensus K-225 is substituted with R-225. Therefore, the ubiquitination site in this location is missing. It seems that this gain of an additional residue has a cumulative effect that could lead to changes in SpBra folding and stability, making it more resistant to ubiquitination enzymes compared to other sea urchin Bra proteins.

Table 3.1 Prediction of ubiquitination sites in four sea urchin species Brachyury protein sequences using UbPred and BDM-PUB software.

Species		Peptide sequence	Position	UbPred score (threshold 0.62)	BDM-PUB score (threshold 0.3)
Strongylocentrotus	purpuratus	EMNRGSE**K**GDPSEKG	33	0.93	1.39
		EKGDPSE**K**GLKVRLD	39	0.98	2.57
		DPSEKGL**K**VRLDDVE	42	0.71	2.00
		GEWIPGG**K**PDGSPPT	113	0.66	Not ubiquitinated
		KQAVNFS**K**VKLTNKL	144	Not ubiquitinated	0.67
		AVNFSKV**K**LTNKLNG	146	Not ubiquitinated	1.98
		SKVKLTN**K**LNGSGQV	150	Not ubiquitinated	1.39
		DITQLKI**K**YNPFAKA	210	Not ubiquitinated	0.63
		AKAFLDI**K**DNRNDGH	222	0.64	1.69
		VHDFQGS**K**YPQFGGW	242	0.62	Not ubiquitinated
		PHHQLIG**K**SVSSWNP	490	0.63	Not ubiquitinated
Hemicentrotus	pulcherrimus	EMNRGSE**K**GDPSEKG	33	0.93	1.39
		EKGDPSE**K**GLKVRLD	39	0.95	2.57
		DPSEKGL**K**VRLDDVE	42	0.70	2.00
		GEWIPGG**K**PDGSPPT	113	0.65	Not ubiquitinated
		KQAVNFS**K**VKLSNKL	144	Not ubiquitinated	1.11

		AVNFSKV**K**LSNKLNG	146	Not ubiquitinated	2.17
		SKVKLSN**K**LNGSGQV	150	Not ubiquitinated	1.80
		DITQLKIK**Y**NPFAKA	210	Not ubiquitinated	0.63
		AKAFLDI**K**DKNDGHD	222	Not ubiquitinated	0.41
		AFLDIKD**K**NDGHDLF	224	0.71	Not ubiquitinated
		VHDFQGS**K**YPQFGGW	241	0.68	Not ubiquitinated
		PPPPYHQ**K**YSAAGAG	297	Not ubiquitinated	2.23
		PHHQLIG**K**SVSSWNP	489	0.67	Not ubiquitinated
Paracentrotus	lividus	EMNRGSE**K**GDPSEKG	33	0.95	1.39
		EKGDPSE**K**GLKVRLE	39	0.98	2.51
		DPSEKGL**K**VRLEDTE	42	0.71	1.89
		GEWVPGS**K**PDGSPPT	113	0.65	Not ubiquitinated
		KQAVNFS**K**VKLSNKL	144	Not ubiquitinated	1.11
		AVNFSKV**K**LSNKLNG	146	Not ubiquitinated	2.17
		SKVKLSN**K**LNGSGQV	150	Not ubiquitinated	1.80
		DITQLKI**K**YNPFAKA	210	Not ubiquitinated	0.63
		AKAFLDI**K**DKNDGHD	222	0.66	0.41
		AFLDIKD**K**NDGHDLF	224	0.67	Not ubiquitinated
		VHDLQGS**K**YPQFGGW	241	0.74	Not ubiquitinated
		PPPPYHQ**K**YAAAGAG	297	Not ubiquitinated	2.19
		PHHQLVG**K**SVSSWNP	489	0.71	0.76
Lytechinus	variegatus	EMNRGSE**K**GDPSEKG	33	0.91	1.39
		EKGDPSE**K**GLKVRLE	39	0.93	2.51
		DPSEKGL**K**VRLEDVE	42	0.71	1.86
		GEWVPGG**K**PDGSPPT	112	0.64	Not ubiquitinated
		KQAVNFS**K**VKLSNKL	143	Not ubiquitinated	1.11
		AVNFSKV**K**LSNKLNG	145	Not ubiquitinated	2.17

SKVKLSN**K**LNGSGQV	149	Not ubiquitinated	1.80
DITQLKI**K**YNPFAKA	209	Not ubiquitinated	0.63
AKAFLDI**K**DKNEGHD	221	0.63	Not ubiquitinated
AFLDIKD**K**NEGHDLF	223	0.64	Not ubiquitinated
VHDLQGS**K**YPQFGGW	240	0.73	Not ubiquitinated
PPPPYHQ**K**YGAAGAG	296	0.52	1.89
PHQQLVG**K**SVSSWNP	489	0.68	0.53

UbPred Legend

Score range	Confidence	Sensitivity	Specificity
0.62 < s < 0.69	Low	0.464	0.903
0.69 < s < 0.84	Medium	0.346	0.950
0.84 < s < 1.00	High	0.197	0.989

Putative SUMOylation and SUMO interaction sites were predicted for four Brachyury sea urchin protein sequences using GPS-sumo and filtered with low stringency.

The predicted results for all four sea urchin Brachyury sequences showed the same SUMOylation and SUMO interaction sites with almost identical scores (Table 3.2).

SUMOylation site was found for K-164, located in the T-box domain, and the SUMO interaction site was detected as the IHIIR sequence located just after the SUMOylation site, in the position 169-173 (168-172 in *L. variegatus*). These results show that SUMOylation pathways are most probably identical for all sea urchin Brachyury proteins, which is highly unlikely to contribute to an increase in SpBra's stability.

Table 3.2 Prediction of SUMOylation and SUMO interaction sites in four sea urchin species Brachyury protein sequences using GPS-sumo software.

Species	Peptide	Position	Score	Cut-off	Type
S. purpuratus	VMLNSLH**K**YEPRIHI	164	35.905	35.288	SUMOylation
S. purpuratus	LHKYEPR **IHIIR** VGGREKQ	169-173	60.32	55.31	SUMO Interaction
H. pulcherrimus	VMLNSLH**K**YEPRIHI	164	35.905	35.288	SUMOylation
H. pulcherrimus	LHKYEPR **IHIIR** VGGREKQ	169-173	60.32	55.31	SUMO Interaction
P. lividus	VMLNSLH**K**YEPRIHI	164	35.905	35.288	SUMOylation
P. lividus	LHKYEPR **IHIIR** VGGREKQ	169-173	60.32	55.31	SUMO Interaction
L. variegatus	VMLNSLH**K**YEPRIHI	163	35.905	35.288	SUMOylation
L. variegatus	LHKYEPR **IHIVR** VGGREKQ	168-172	60.196	55.31	SUMO Interaction

Stringency Legend

SUMOylation	SUMO Interaction
Low	**Low**
Medium	Medium
High	High

This observation, however, does not exclude the importance of SUMOylation in Brachyury. As previously mentioned, *C. elegans* TBX-2 and human TBX22 TFs are modified by SUMOylation that allows them to act as repressors (Roy Chowdhuri *et al.*, 2006; Andreou *et al.*, 2007; Crum and Okkema, 2007). Even if the Crum and Okkema study from 2007 points out that SUMOylation *site-1*, located at the C-terminal portion of the T-box domain and conserved in many T-box TFs, does not appear in Brachyury, it could be possible that sea urchin Brachyury has another site

located in the T-box domain that might affect its ability to act as a repressor. This novel site could be the one predicted in this thesis work.

3.2.4 The interplay between Brachyury and Smad: what we can learn from sea urchins

Interaction between Brachyury and Smad1 TFs and its possible synergistic involvement in metazoan development was first described more than a decade ago by the Smith group (Messenger *et al.*, 2005; Marcellini, 2006). The earlier studies highlighted Brachyury proteins' inductive abilities to be determined by the N-terminal domain and not the T-box domain (Marcellini *et al.*, 2003). First described in *Xenopus*, the interaction between Brachyury and Smad1 TF occurs, as described previously, through a consensus sequence HLLXAVX. This sequence appeared most probably, very early in the metazoan evolution, since it was found in the genomes of different protostomes: mollusks, annelids, chaetognaths, insects (secondarily lost in dipterans), and most deuterostomes (the sequence was lost in urochordates). All echinoderm Brachyury protein sequences show a very high conservation of this protein region (Figure 3.9).

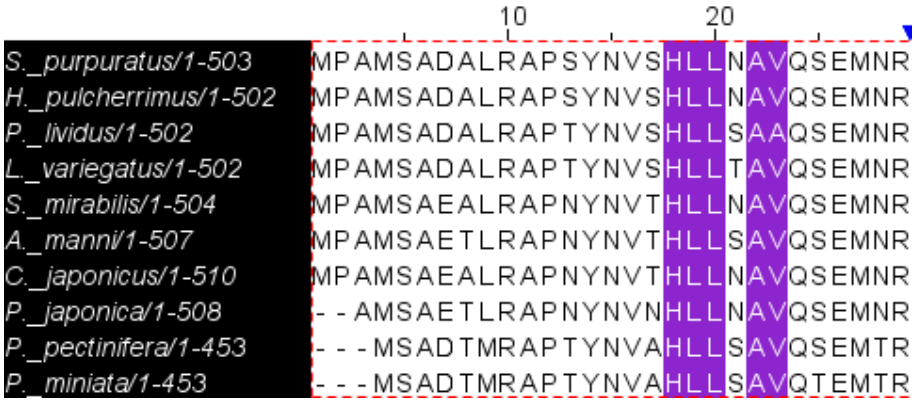


Figure 3.9 N-terminal domain of different echinoderm Brachyury proteins with highlighted Smad1 interacting consensus domain. Sea urchins (*S. purpuratus*, *H. pulcherrimus, P. lividus, L. variegatus)*; sand dollars (*S. mirabilis, A. manni, C. japonicus, P. japonica*), and starfish (*P. pectinifera*, *P. miniata)*

Some later studies have shown that Brachyury is able to interact with Smad2/3 TF, most probably not directly, but as a cofactor of Eomesodermin that could directly bind to Smad2/3 (Faial *et al.*, 2015). The Eomes orthologue in echinoderms is T-Brain (Tbr), which is a known mesodermal (skeletogenic) marker. If we consider Tbr – Smad2/3 interaction as ancestral and conserved, it is unlikely that Bra could interact with Smad2/3 since Bra and Tbr do not share the same spatial domain of expression in echinoderms unless there is another, unknown mechanism that allows that interaction. To confirm that Bra-Smad1 interaction is conserved among echinoderms, the protein sequenced was searched in existing databases of sea urchins *P. lividus* and *L. variegatus*, sea cucumber *P. parvimensis*, and starfish *P. miniata*. Smad1/5/8 (probably an ancestral TF that evolved in later deuterostomes into separate Smad1, Smad5, and Smad8) sequences were aligned and checked for the presence of consensus Brachyury interacting sites in the MH2 domain. Each species had multiple predicted proteins based on the transcriptome data, and based on the alignment results, *P. lividus*, *P. miniata*, and *P. parvimensis* Smad1/5/8 had the consensus sequence, while *L. variegatus* seem to have lost it (Figure 3.10). One explanation could be that the *L. variegatus* sequencing data was incomplete. However, since there is a lack of available data for other echinoderm species, no firm conclusions can be made.

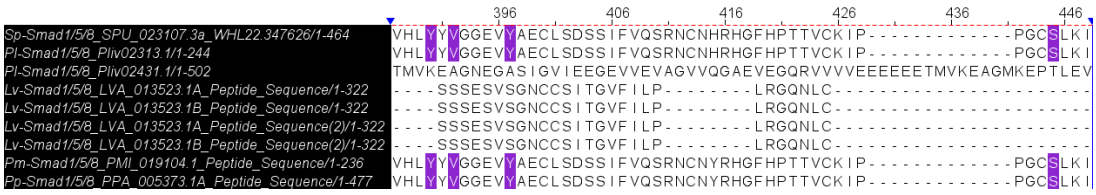


Figure 3.10 Multiple echinoderm Smad1/5/8 MH2 domain sequence alignment and the presence of Brachyury interacting consensus sites. Sp: *Strongylocentrotus purpuratus*, Pl: *Paracentrotus lividus*, Lv: *Lytechinus variegatus*, Pm: *Patiria miniata*, Pp: *Parastichopus parvimensis*. The consensus sites are marked. *P. lividus* peptide Pliv02431.1 and all *L. variegatus* peptides do not possess the Brachyury interacting sites in their MH2 domains.

Even though much information is still missing, one could be certain that Brachyury-Smad1/5/8 interaction exists, at least in some echinoderms. Moreover, since this

mechanism of interaction seems to be ancestral, Brachyury may be able to interact with other Smad orthologues. In addition, the interaction with Smad protein could be responsible for activating or repressive action of Brachyury during echinoderm development.

3.3 Conclusions

Different hypotheses could explain the Brachyury protein presence in the archenteron of the late gastrula in *S. purpuratus*. It may be possible that the *SpBra* mRNA is present in the gut cells, and it allows the synthesis of the SpBra protein. However, due to its low level of expression, it cannot be detected by the conventional *in situ* hybridization. Or, the mRNA is not present in the gut cells anymore; however, the presence of the protein could be explained by its particularly high stability. Another explanation could be an antibody sensitivity issue. It may be possible that the SpBra protein detection in the archenteron is the detection of inactive protein that starts to be targeted by ubiquitination mechanisms for degradation.

Even though the experimental part was somewhat inconclusive, the computational analyses gave some promising results.

The protein sequence-structure analysis and secondary structure prediction showed some significant differences between the echinoderm Brachyury proteins. First, the primary structure analysis had revealed important differences between the SpBra protein compared to other echinoderms. Particularly important findings are in the C-terminal domain: the acquirement of an extra asparagine residue and the loss of lysine in the same region, adjacent to the T-box domain. These changes could have importance in stabilizing the protein since one ubiquitination site could be missing. Moreover, other changes in the C-terminal domain, in the regions spanning from 381-391 and 411-421, SpBra had acquired proline, alanine, and leucine. These changes could have made an evolutionary impact on the formation of helical structures that could have affected its folding, stability, and interaction with other molecules. However, considering the fact that the secondary structure prediction is

based on known crystallographic structures, only the conserved DNA (T-box) binding domain structure prediction can be considered highly significant based on the published structures of *Xenopus* Brachyury protein.

Looking for Smad1 orthologues in echinoderm genomes and the presence of Smad1 interaction consensus sequence in different echinoderm Brachyury proteins suggests that Brachyury-Smad1 interaction is indeed ancestral and that it should be studied more as an important synergistic transcription factor mechanism of action in development.

Finally, the differences between the various sea urchin Brachyury proteins could be explored by additional analyses in the future, for example, using different biochemical assays to measure the protein stability or expressing the SpBra protein in other species utilizing transgenesis, such as performing a CRISPR-Cas9 knock-in of the *SpBra* DNA sequence into *P. lividus*.

Contribution statement

Whole-mount immunofluorescence experiments were performed by Periklis Paganos, a PhD student from Dr. Arnone's lab, and the author of this thesis. The author of this thesis performed all *in silico* analyses described in this chapter.

CHAPTER 4

DIFFERENTIAL GENE EXPRESSION ANALYSIS AFTER BRACHYURY KNOCK-DOWN IN *STRONGYLOCENTROTUS PURPURATUS*

This chapter discusses the results arising from the Brachyury knock-down experiment in the sea urchin species *Strongylocentrotus purpuratus*. The combination of various datasets (differential RNA-seq, Brachyury ChIP-seq, ATAC-seq, and fluorescent *in situ* hybridization) was used to uncover the putative indirect and direct Brachyury targets. Moreover, the updated GRNs around Brachyury at various developmental stages are presented in this thesis chapter.

4.1 Introduction

Reconstructing a GRN around a specific transcription factor (TF) requires different points of view – spatial and temporal expression of a TF of interest, the developmental process in which this TF might be involved, the interaction of this TF with other molecules, including both DNA and signaling molecules. So far, the use of perturbation techniques, especially morpholino antisense oligonucleotide (MASO) (Summerton, 1999) injection against the gene of interest has been widely used to discover possible nodes of the GRN. Monitoring the morphological changes and gene expression changes after a perturbation has been widely used in the evo-devo field (Yamada *et al.*, 2003; Mende, Christophorou and Streit, 2008; Timme-Laragy, Karchner and Hahn, 2012; Materna, 2017).

Differential transcriptomic approach, after the perturbation of a given TF of interest using a MASO, gives a vast amount of information regarding the downregulation or upregulation of genes, such as certain transcription factors and signaling molecules.

However, it does not give any answer to whether the detected changes might be under the direct effect of the perturbed TF or not. Therefore, other omics approaches, in addition to differential RNA-seq, should be combined to achieve this goal. Systems biology approaches look at all single elements of a system and their interactions, and in this case, a gene regulatory state could be considered a unique, separate system. The GRN system is composed of different interactions among transcription factors, signaling molecules, and terminal differentiation genes, and therefore, must be observed as a whole. The putative downstream genes from a transcription factor of interest can be analyzed by assessing the effect of this TF on the target and looking at the location of the target in the genome and the DNA structure upstream and downstream of putative targets, and whether those DNA regions are regulatory. Luckily, nowadays, we are equipped with powerful methods that allow us to study the transcription factors of interest and their effect on the target via a systems biology approach.

In this chapter, this integrative approach of combining different omics techniques to detangle the role of Brachyury during the gastrulation of *S. purpuratus* is described.

4.1.1 The use of RNA-sequencing for precise transcript quantification after perturbation

Transcriptome represents a snapshot in time of a cell RNA state (Brown TA, 2002). RNA-seq is a technique that uses high-throughput sequencing methods in detecting RNA levels of a cell, tissue, or embryo. However, RNA levels are not measured directly; RNA is first retro-transcribed into cDNA, amplified, and then sequenced. The advantage of Next Generation Sequencing (NGS) technologies in assessing RNA quantification is producing millions of short reads in a relatively short time (Kukurba and Montgomery, 2015). The most common use of RNA-seq is comparing gene expression levels between two or more conditions, also known as differential RNA-seq. Differential RNA-seq is one of the most important methods in assessing gene expression levels during development, such as comparison of different developmental stages, wild-type versus drug treatment, or comparison of normal and perturbed gene conditions (Lowe, Cuomo and Arnone, 2016, 2017). In this

work, the advantage of the bulk RNA-seq technique is used to assess the effects of Brachyury absence on the mRNAs involved in the gastrulation process of the *S. purpuratus.*

4.1.2 Chromatin immunoprecipitation sequencing – discovering genes under direct influence of a transcription factor

Chromatin immunoprecipitation (ChIP) is a technique that allows *in vivo* assessment of the interaction between the proteins and their binding sites on the DNA molecule (Collas, 2010). Chromatin represents the complex of DNA packed with histone proteins (Hübner, Eckersley-Maslin and Spector, 2013). The regions of accessible, or so-called open chromatin represent possible active sites where regulatory proteins, such as transcription factors, can bind to modulate the chromatin organization or drive the expression of certain genes (Hübner, Eckersley-Maslin and Spector, 2013). ChIP technique uses the physically sheared chromatin, in which the protein-DNA interactions are preserved by formaldehyde fixation to selectively precipitate the protein-bound DNA fragments using a specific antibody (Solomon, Larsen and Varshavsky, 1988; Orlando, 2000). The precipitated fragments can be analyzed with different methods, like qPCR, or with high throughput sequencing. ChIP-seq combines the Chromatin immunoprecipitation method with high throughput sequencing (Barski *et al.*, 2007). ChIP-seq method is used for assessing the chromatin organization (Adli and Bernstein, 2011; Song and Smith, 2011), or the gene regulation driven by protein-DNA interactions (Lei *et al.*, 2010). It is based on sequencing specific protein-bound DNA fragments that are mapped to the genome after the sequencing reads are generated. In developmental biology, the ChIP-seq method is applied to discover locations of transcriptional complexes and detection of enhancers (Visel *et al.*, 2009) or to identify Cis-Regulatory modules, which can help in improving already known or drafting new GRNs (Lindeman *et al.*, 2009; Khor, Guerrero-Santoro and Ettensohn, 2019). In this work, this technique's advantage is used to discover the chromatin regions that show enrichment in fragments where Brachyury TF is bound at two developmental stages, the 24hpf mesenchyme blastula and the 48hpf late gastrula.

4.1.3 Assay for Transposase-Accessible Chromatin using sequencing in evo-devo

Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) is a novel technique developed by Buenrostro and colleagues in 2013 (Buenrostro *et al.*, 2013). ATAC-seq method uses a hyperactive Tn5 transposase enzyme that can cut the open, histone-free chromatin regions and add sequencing adapters (Buenrostro *et al.*, 2013). The DNA fragments that were cut and tagged with sequencing adapters are amplified and sequenced using the high-throughput sequencing technology (Buenrostro *et al.*, 2015). The sequenced reads are then mapped to the genome. ATAC-seq allows finding the chromatin regions accessible to the binding of specific regulatory molecules and transcription factors.

The open chromatin regions discovered by ATAC-seq were used combined with ChIP-seq detected regions of Brachyury TF binding to find the most probable direct Brachyury targets. In this work, all the mentioned techniques were used in a combined manner to detangle the role of Brachyury during the *S. purpuratus* early development. The analysis and the comparison of four RNA-seq datasets are described. The datasets were obtained by Brachyury perturbation at two different developmental stages, the early gastrula and the late gastrula stage of the sea urchin species, *Stronglylocentrotus purpuratus*. The perturbation was accomplished by knocking down Brachyury protein injecting the specific antisense morpholino oligonucleotides into the sea urchin zygotes. Furthermore, the combination of differentially expressed genes discovered by the differential transcriptomic approach combined with the Brachyury targets discovered by ChIP-seq, linked to the open chromatin regions discovered by analyzing ATAC-seq datasets, which was used to discover the putative direct targets of Brachyury, is presented. Finally, the Gene Regulatory network models around Brachyury at two developmental stages are proposed.

4.2 Results and Discussion

4.2.1 From Brachyury knock-down to differentially expressed gene analysis

The choice of developmental stages was based on the pattern of expression of the Brachyury transcription factor. Its first signs of expression are seen at the blastula stage, at 15hpf in *S. purpuratus,* first in the vegetal pole of the embryo, in the territory of the ring of cells marked as veg2 cells, which are progenitors of endodermal and mesodermal tissues (Croce, Lhomond and Gache, 2001; Gross and McClay, 2001; Rast *et al.*, 2002; de-Leon and Davidson, 2010; Peter and Davidson, 2011a). This vegetal ring is composed of cells described as endomesoderm territory made up of two rings - the inner ring that will give rise to mesoderm and the outer ring to anterior endoderm. Later on, from 24hpf, at the late blastula stage, *brachyury* expression is switched off in the veg2 lineage, and the expression starts in the surrounding ring of veg1 cells. This circular region will become known as the blastopore when the gastrulation starts after 24hpf, and this is the region where *brachyury* expression will be maintained during later larval stages, after the veg1 cells will be invaginated as well, to become the posterior part of the gut. Apart from the vegetal pole of the embryo, *brachyury* is expressed in the oral ectoderm region starting from 21hpf (mesenchyme blastula), and it is maintained in this domain throughout the development as a gene involved in stomodeum (mouth) formation (Rast *et al.*, 2002). The same expression pattern is detected in all sea urchins tested so far – *P. livuds, L. variegatus, and A. lixula* (Croce, Lhomond and Gache, 2001; Gross and McClay, 2001; Andrikou, 2012). To understand more deeply the involvement of Brachyury in regulating endoderm and ectoderm development and patterning of the gut, two developmental stages were selected – the 27hpf early gastrula and the 48hpf late gastrula. At 27hpf, the blastopore is already open, and *brachyury* is already highly expressed in both domains. At 48hpf, the gastrulation is finished, and, as described in Chapter 3, Brachyury protein localization has been observed in the gut of the late gastrula embryo, apart from the oral ectoderm and the area around the blastopore.

Differential gene expression analysis. The PCA plots show that control and treated samples coming from the same batch of embryos cluster close to each other (Figure 4.1). That means the subject effects are larger than the treatment effects (Love, Huber and Anders, 2014). Therefore, the batch effects were removed before running the differential expression analysis.

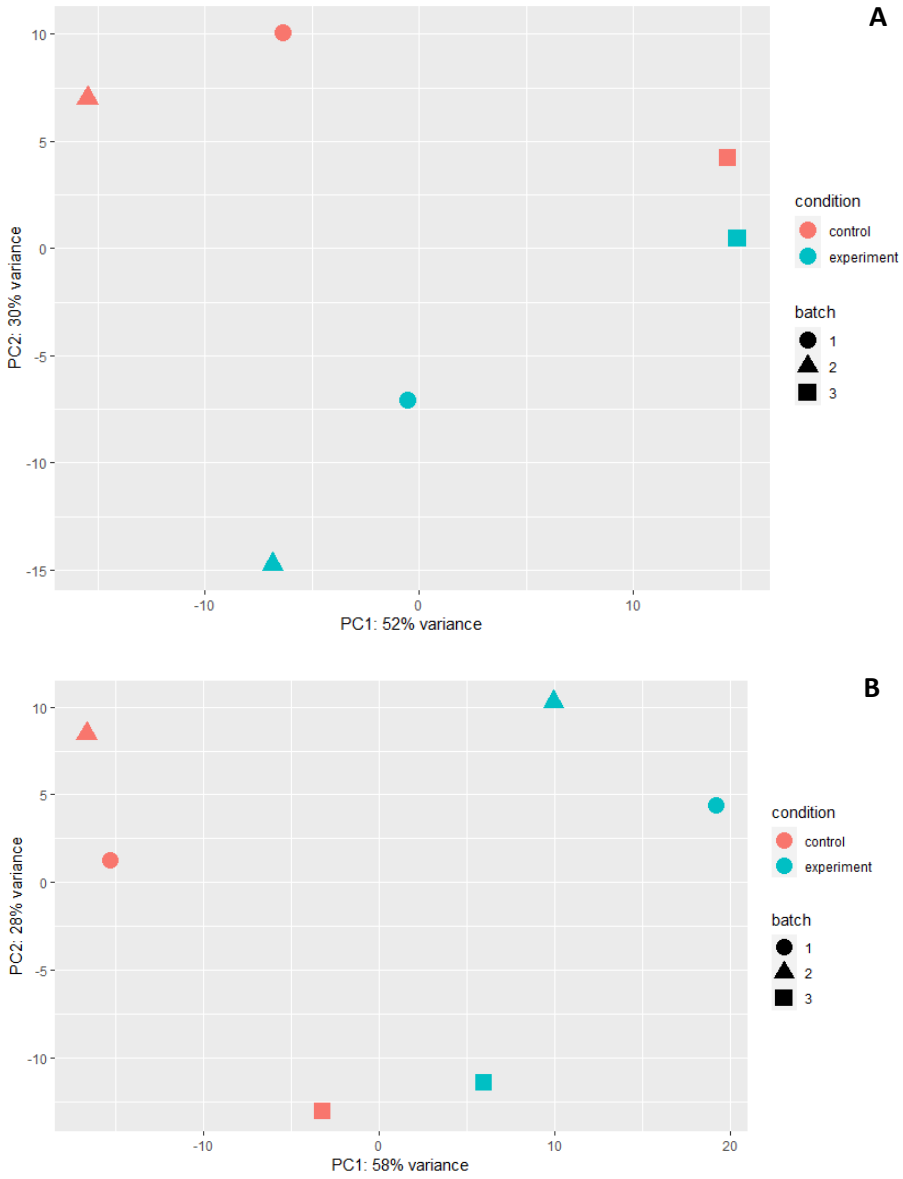


Figure 4.1 Principal component analysis (PCA) of the 27hpf and 48hpf *S. purpuratus* Brachyury perturbed embryos. A) Sequenced samples at 27hpf; B) Sequenced samples at 48hpf. Different colors represent the experimental condition, while different shapes represent different batches PCA plots show close clustering of samples that come from the same batches of embryos.

To further characterize the direct and indirect Brachyury targets at the early gastrula and late gastrula stages, the differently expressed genes were coupled with the ChIP-seq/ATAC-seq datasets. *De novo* motif discovery was performed, as described in Chapter 2, in order to discover which motifs sequences are the most enriched in the ChIP-seq datasets.

To gain insight into which transcription factors and signaling molecules are under the direct influence of Brachyury, the differentially expressed transcription factors and signaling molecules detected after the knock-down of Brachyury were checked for presence in the ChIP-seq datasets.

4.2.2 Phenotype analysis

To ensure that the MASO injection was successful and to make a biological sense of generated omics data, it is quite important to assess the morphological differences between the wild type and perturbed samples. To this aim, the wild type and Brachyury knock-down embryos at 27 and 48hpf were analyzed for any morphological differences.

In the wild type condition at the early gastrula (27hpf) stage, the embryo has a form of a hollow ball with a thicker flattened vegetal plate that starts to change its shape around its center, as described in the first chapter. The veg2 lineage already started to invaginate to give rise to the endoderm - archenteron, and the blastopore is already open (marked in Figure 4.2 in the upper left corner). All the primary mesenchyme cells (PMCs), which will later become skeletal cells, have already ingressed inside the blastocoel (Arrow in the upper left corner of Figure 4.2). On the contrary, Brachyury MASO injected embryos show delay in development since the blastopore cannot be observed. Moreover, a higher number of PMCs can be observed (Figure 4.2. upper right corner).

Wild type late gastrula shape has the already completely formed tube-like archenteron bent towards the ventral (oral) ectoderm of the embryo, and it starts to adopt the prismatic shape (Figure 4.2, lower left corner, red arrow). PMCs are re-arranged in the embryo's vegetal pole forming a circle where the skeletogenic

spicules will form. Again, the Brachyury MASO injected embryo has a visible delay in development. The ectodermal patterning seems to be affected since the embryo shape is still round. The archenteron shows some malformations in its shape – it does not form the straight tube, and despite the fact that the gut orientation is still towards the stomodeum, the characteristic shape of ventral and dorsal ectoderm, flattened (Figure 4.2 red arrows) and bulging (Figure 4.2 green arrows) respectively, is not visible. Moreover, it seems that the blastopore opening has shifted its position towards the ventral (oral ectoderm) region.

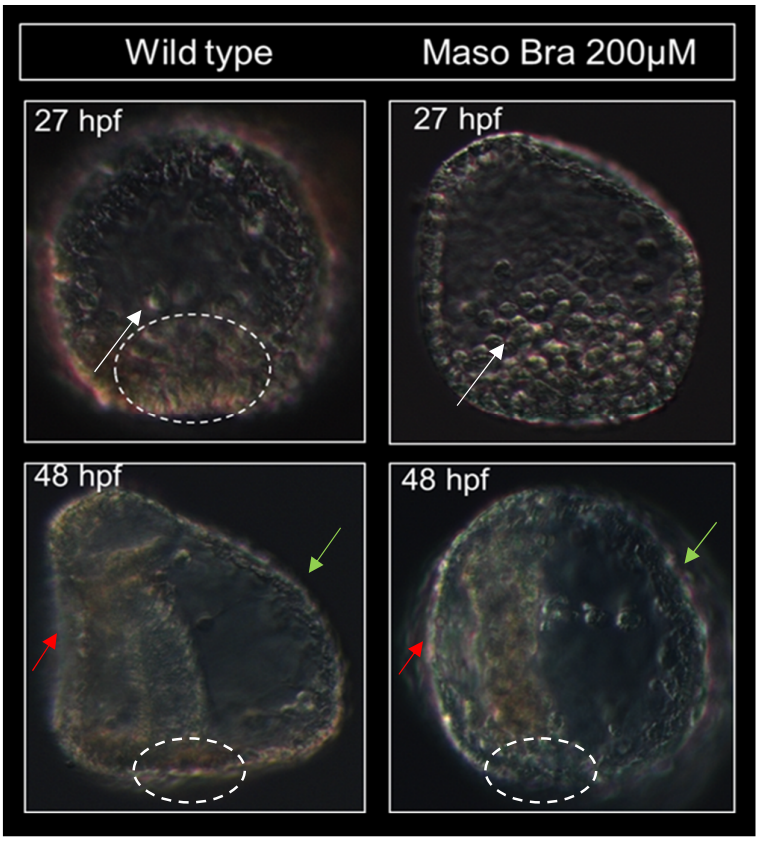


Figure 4.2 Phenotypic differences in control and Brachyury MASO treated embryos at early gastrula (27hpf) and late gastrula (48hpf) of *S. purpuratus*. The white dotted oval shapes represent the blastopore opening. The white arrows in the 27hpf control and treated embryos represent the PMCs. 48hpf embryos are represented in the lateral view. Red arrows represent the ventral side, while the green arrows represent the dorsal side of the embryo. Courtesy of Periklis Paganos.

In conclusion, the phenotypic differences show the effects of the Brachyury protein absence on all cell lineages – ectodermal, visible in the overall embryo shape in the late gastrula which does not start to form a prismatic shape; mesodermal, visible in the higher number of primary mesenchyme cells in the early gastrula stage of injected embryo compared to the wild type; and endodermal visible in both stages, the blastopore opens later and it is moved towards the oral ectoderm, and the gut shows a somewhat irregular shape compared to the wild type.

4.2.3 Gene ontology enrichment analysis of differentially expressed genes in S. purpuratus *of 27hpf and 48hpf DEG*

Gene ontology enrichment analysis of the differentially expressed genes was conducted to discover the class of proteins and biological processes affected by Brachyury misexpression.

At the early gastrula stage (27hpf), there were 279 differentially expressed genes detected after the knock-down of Brachyury, while at late gastrula (48hpf), the number was 1087 genes. At the early gastrula, 111 genes were detected as downregulated (40%), and 168 as upregulated (60%); while at the late gastrula, 646 were downregulated (59%) and 441 upregulated (41%).

For the 27hpf gene data set, 253 genes out of 297 were tested for the gene ontology enrichment terms, while for 48hpf, 915 genes out of 1088 were tested to detect which biological processes and protein classes were affected by the absence of Brachyury. The functional terms for the *S. purpuratus* genome are partially manually and partially electronically annotated, and many genes have an utterly unknown function (Kudtarkar and Cameron, 2017). For this reason, the number of hits for each category is lower than the input. The gene lists were classified for biological processes and protein class terms.

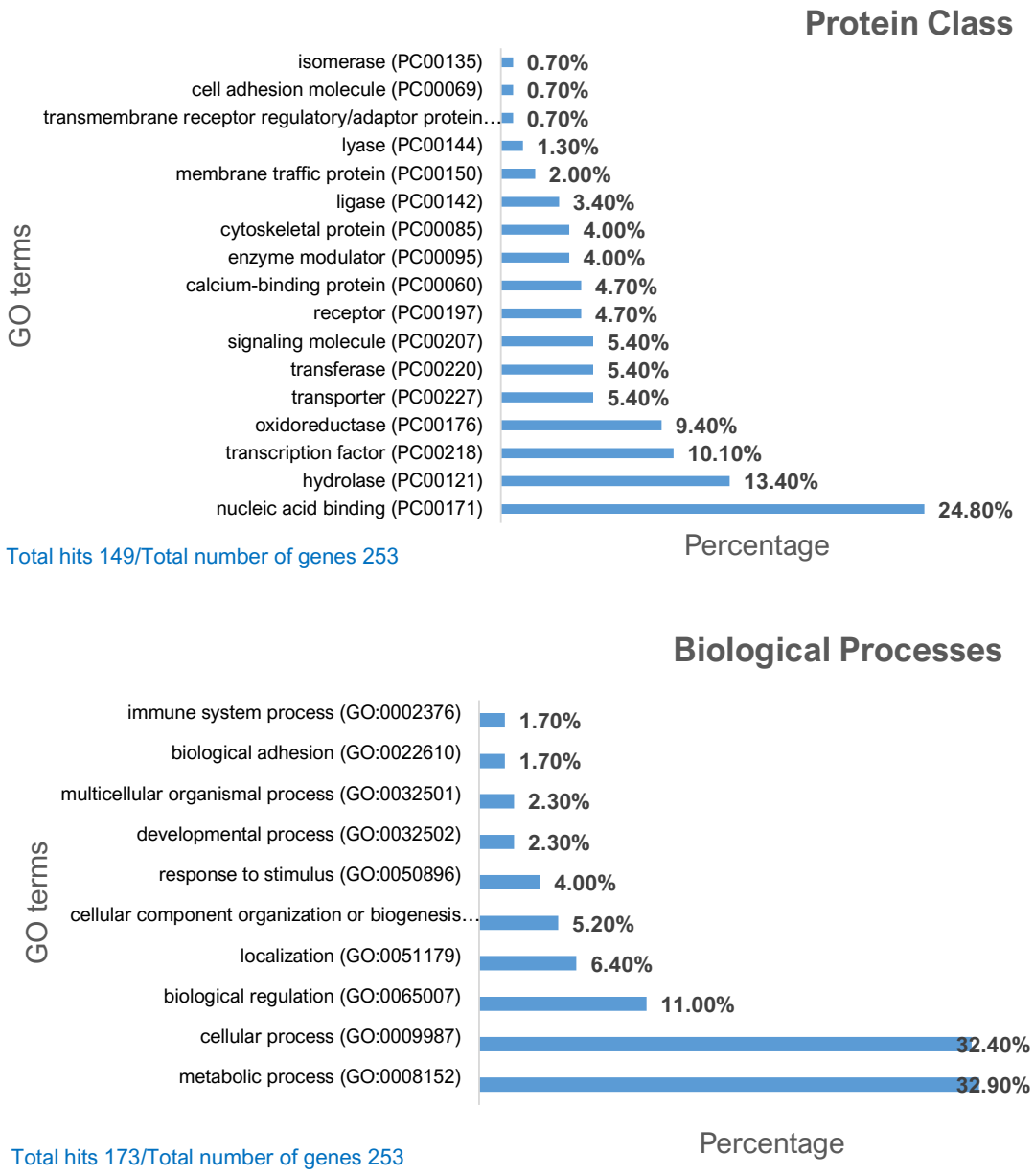
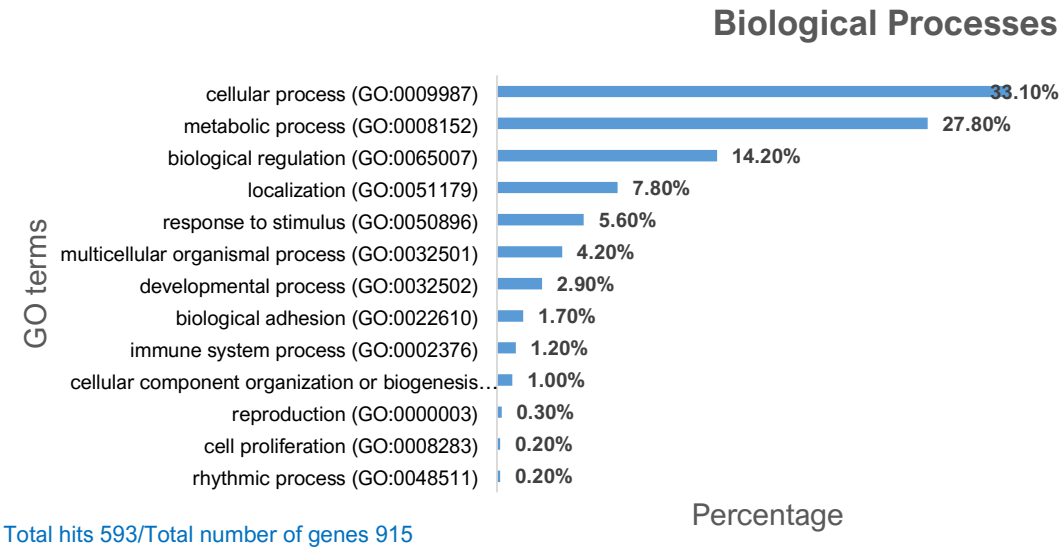


Figure 4.3 Gene Ontology terms analysis for 27hpf *S. purpuratus* Brachyury knock-down embryos. Each graph represents a specific category and contains specific GO terms and the percentage of genes associated with them. Lower left corner of each graphs shows the number of analyzed genes compared to the total number of differentially expressed genes. The upper graph represents the Biological Processes and the lower graph represents Protein class affected by the absence of Brachyury.

At both early and late gastrula, the most highly affected Biological processes after the knock-down of Brachyury were metabolic and cellular. The annotation of the Gene Ontology terms related to the metabolism came about shortly after the *S. purpuratus* genome sequencing, and it showed that 2300 proteins were found to be related to metabolic transport and the enzymatic conversion (Goel and Mushegian, 2006). Looking closely and the exact number of differentially expressed genes connected to metabolism activity showed that, at the early gastrula, 23 genes were associated with metabolic processes, precisely to energy, amino acid, carbohydrate, lipid metabolism, and protein degradation. In contrast, at 48hpf, the number increased to 104 (For more details, refer to the Non-book component). This can point to the importance of Brachyury in regulating the cell and embryo growth. The study by Marsh *et al*. showed that metabolism-related processes before the late larval stages in sea urchin are involved in harvesting the maternally present nutrients to promote the cell proliferation, leading to the embryo and larval growth (Marsh, Leong and Manahan, 1999). The number of genes involved in metabolism increases proportionally to the number of cells present in the embryo or larva (Marsh, Leong and Manahan, 1999). Therefore, from gene ontology data gathered in this work, it could be inferred that Brachyury is one of the important TFs that regulate the growth of the embryo based on the regulation of the genes involved in the metabolism of the maternally present sources of energy.

The top 10 protein classes most affected by Brachyury's absence at both developmental stages were nucleic acid binding, hydrolase, transcription factor, oxidoreductase, transporter, transferase, signaling molecule, receptor, calcium-binding, and enzyme modulator (Figure 4.3B and Figure 4.4B). The high number of hydrolases, oxidoreductases, transporters, transferases, and enzyme modulators confirm Brachyury's important role in metabolism regulation. A high number of affected nucleic acid binding proteins, transcription factors, signaling molecules, and receptors shows that Brachyury has an important role in the early regulation of basic gene regulation activity associated with embryo development.



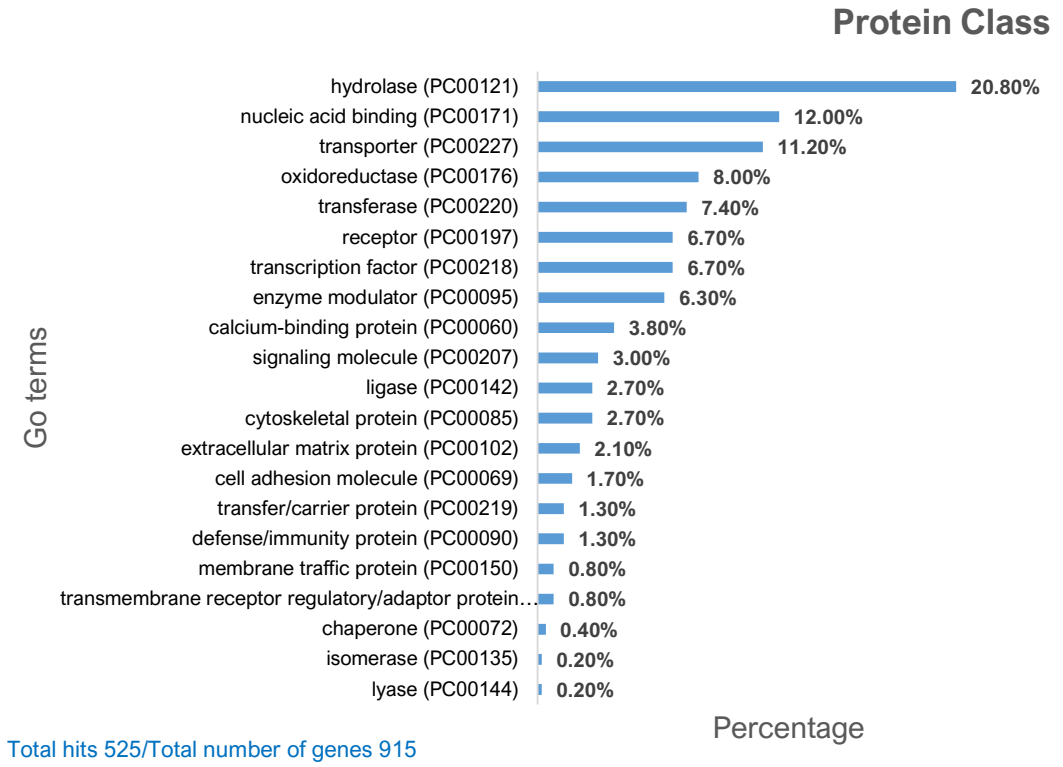


Figure 4.4. Gene Ontology terms analysis for 48hpf *S. purpuratus* Brachyury knock-down embryos. Each graph represents a specific category and contains specific GO terms and the percentage of genes associated with them. Lower left corner of each graphs shows the number of analyzed genes compared to the total number of differentially expressed genes. The upper graph represents the Biological Processes and the lower graph represents Protein class affected by the absence of Brachyury.

Brachyury is an important upstream regulator of different events leading to defining the development of different embryo domains, and it is explored in detail in the next subsections of this Chapter.

4.2.4 Differentially expressed transcription factors and signaling molecules after Brachyury knock-down

To untangle the role of Brachyury TF and to refine and update the currently known GRNs, transcription factors and signaling molecules affected by its perturbation were investigated. Moreover, since the high number of differentially expressed genes after the depletion of Brachyury TF belong to the protein classes involved in transcriptional processes, this work's main focus was to investigate them in more detail.

Transcription factors and signaling molecules affected at 27hpf

At 27hpf, in total, 35 TFs/signaling molecules were affected, from which 12 (34%) were downregulated, and 23 (66%) were upregulated (Table 4.1).

Table 4.1 Differentially expressed transcription factors and signaling molecules after Brachyury knock-down at 27hpf in *S. purpuratus*. Each gene is described by its SPU id number, gene name, logarithmic fold change (log2FC), spatial and temporal expression pattern and a reference. The list of perturbed genes is sorted based on the intensity of perturbation starting with the highest downregulation (the darkest blue color) and ending with the highest upregulation (the darkest red color).

SPU ID	Gene name	Gene description	log2FC	Spatial expression	Reference
SPU_020472	Sp-Hypp_975	hypothetical protein-975	-2.244	unknown	n.a.
SPU_023730	Sp-Isl	insulin gene enhancer protein	-1.684	Blastula oral ectoderm; Glastula and pluteus: CB, foregut, anus	Perillo et al., 2018; Annunziata et al., 2013
SPU_019089	Sp-Z133	zinc finger protein 133	-1.258	Blastula: apical ectoderm; Glastula: apical ectoderm	Materna et al., 2006

SPU_012491	Sp-Nk1	NK1 homeobox	-1.065	Blastula: veg1 ecto (oral ectoderm border with endoderm); Glastula: blastopore/vegetal plate	Minokawa et al., 2004
SPU_003591	Sp-Nan2	Nanos2	-0.933	Blastula: small micromeres; Glastula: tip of the archenteron	Luo & Su, 2012; Juliano et al., 2006
SPU_002634	Sp-Hox7	homeobox 7	-0.869	aboral ectoderm	Howard-Ashby et al., 2006
SPU_015767	Sp-Z57	zinc finger protein 57	-0.783	unknown	Materna et al., 2006; Tu et al., 2012; Materna et al., 2013; Cui et al., 2014
SPU_011837	Sp-Apobec1	Apolipoprotein B mRNA editing enzyme	-0.752	Blastula: vegetal plate (veg2 meso and veg2 endo), Glastula: gut, SMC	Materna & Davidson, 2012; Peter & Davidson 2011
SPU_011271	Sp-Sfrp1/5	secreted Frizzled related protein 1/5	-0.724	Blastula: apical ectoderm	Range & Wei, 2017
SPU_007452	Sp-Spec1	Spec 1a	-0.589	Blastula: aboral ectoderm	Hardin et al., 1985
SPU_010351	Sp-IrxA	iroquois homeobox A	-0.584	Blastula: aboral ectoderm, veg1 ecto, Early glastula: oral ectododerm	Howard-Ashby et al., 2006; Su et al., 2009; de-Leon et al., 2013
SPU_006676	Sp-FoxA	forkhead box A	-0.517	Blastula: oral ectdoerm, vegetal plate, Glastula: gut, stomodeum	Oliveri et al., 2006
SPU_004717	Sp-Spry/Socscp	Sprouty homolog	0.494	unknown	n.a.
SPU_015984	Sp-Prox1	prospero-related homeobox 1-like	0.577	Blastula: veg2 meso; Glastula: tip of the archenteron	Materna & Davidson, 2012
SPU_014793	Sp-Spalt	sal-like	0.640	Blastula: vegetal plate - micromeres and veg2 meso; Glastula: gut	Materna et al., 2006
SPU_024903	Sp-Ese	epithelium-specific ets factor-like	0.698	Blastula: veg2 meso; Glastula: SMC	Rizzo et al, 2006

SPU_002874	Sp-Ets1/2	v-ets erythroblastosis virus E26 oncogene homolog 1/2-like	0.729	Blastula: veg2 meso, PMC; Glastula: SMC, PMC	Rizzo et al., 2006
SPU_012469	Sp-Elk	Elk1/3/4-like	0.757	Blastula: vegetal plate; gastrula: ubiq	Rizzo et al., 2006
SPU_006917	Sp-Runt1	runt-related transcription factor-like-1	0.775	Blastula: vegetal plate, PMC, oral ecto; Glastula: gut, stomdoeum, SMC	McCarty & Coffman, 2013; Robertson et al., 2008
SPU_012448	Sp-Ovo	ovo-like, z157	0.789	Blastula: PMC; Glastula: no detection	Materna et al., 2006; Juliano et al., 2006
SPU_016128	Sp-Delta	Delta (Dl) homolog	0.799	Blastula: veg2 meso	Sweet et al., 2002; Sharma & Ettensohn, 2010; Croce & McClay, 2010; Peter & Davidson
SPU_028093	Sp-Scl	stem cell protein-like	0.812	Blastula: veg2 meso	Solek et al., 2013
SPU_025302	Sp-Alx1	aristaless-like homeobox 1-like	0.815	Blastula:and Glastula: PMC	Ettenson et al., 2003; Howard-Ashby et al., 2006
SPU_007742	Sp-Egf/Ig/Lnb/Tm7/Gpcr	Egf/Ig/Lnb/Tm7/Gpcr	0.827	unknown	n.a.
SPU_020451	Sp-Bra	Brachyury	0.859	Blastula: veg endo, oral ectoderm; Glastula: blastopore, hindgut, stomodeum	Gross & McClay, 2001; Croce et al., 2001; Rast et al., 2002
SPU_021172	Sp-Fra2	Fos(FBJ osteosarcoma oncogene)-like	0.874	Blastula: PMC; Glastula: PMC	Russo et al., 2014; Russo et al., 2018
SPU_006353	Sp-Pde4Bi1	cAMP-specific phosphodiesterase isoform PDE4B1	0.886	unknown	n.a.
SPU_027015	Sp-GataC	GATA binding protein C	0.934	Blastula: SMC; Glastula: SMC, PMC	Davidson et al., 2002; Solek et al., 2013
SPU_015358	Sp-Egr	early growth response, z60	0.936	Blastula: oral and aboral ectoderm; Glastula: apical organ	Materna et al. 2006; Slota et al., 2019

SPU_013178	Sp-Nr1m3	nuclear receptor subfamily 1 group M member 3, NR1-like	0.975	unknown	n.a.
SPU_003102	Sp-Jun	jun proto-oncogene-like, AP-1-like	1.046	Blastula: PMC; Glastula: PMC	Russo et al., 2014; Dylus et al., 2016; Russo et al., 2018
SPU_008112	Sp-Phb1	paired homeodomain 1	1.084	unknown	n.a.
SPU_006032	Sp-Hypp_1626	hypothetical protein-1626	1.227	unknown	n.a.
SPU_026196	SPU_026196	unknown	1.717	unknown	n.a.
SPU_021119	Sp-MyoD1	myogenic differentiation 1-like	2.872	Blastula:and Glastula: PMC	Beach et al., 1999; Andrikou et al., 2013

Based on the known spatial expression patterns, 27 genes were ectodermal, 15 mesodermal, 3 endodermal, and 1 ubiquitously expressed.

In total, 6 oral, 4 aboral, and 2 apical ectoderm TFs/signaling molecules were affected by the absence of Brachyury. Considering mesodermally expressed TFs/signaling molecules, 9 are expressed in the SMC, 6 in the PMC, and 1 in the micromere descendants in the normal developing embryos. Only three genes are expressed in the endodermal linage in the vegetal plate in the normal developing embryos (Table 4.1).

Based on the numbers of significantly affected genes, and the log2fold change, it could be speculated that Brachyury is deeply involved in the regulation of ectodermal patterning and mostly acts as an ectodermal activator. It seems that, based on the high number of upregulated mesodermal genes, Brachyury could be involved in the repression of mesodermal genes; and finally, only a few endodermal genes seem to be activated by Brachyury.

It can be concluded that Brachyury affects all three germ layers, ectoderm, and endoderm at the onset of gastrulation.

Transcription factors and signaling molecules affected at 48hpf

At 48hpf, the number of affected genes had increased 4 times, which is also evident in the subset of TFs/signaling molecules. In total, 97 genes belonging to these categories were affected, from which 54 (56%) were downregulated, and 43 were upregulated (44%) (Table 4.2). This suggests that during development, the effect of Brachyury protein knock-down becomes larger. This can be explained partly due to a higher level of genes being expressed as development progresses, and also, Brachyury having a larger effect in the developing gut. Moreover, the cascade of events happening during gastrulation, when the Brachyury is not expressed, could lead to this very high number of differentially expressed genes.

Table 4.2. Differentially expressed transcription factors and signaling molecules after Brachyury knock-down at 48hpf in *S. purpuratus*. Each gene is described by its SPU id number, gene name, logarithmic fold change (log2FC), spatial and temporal expression pattern and a reference. The list of perturbed genes is sorted based on the intensity of perturbation starting with the highest downregulation (the darkest blue color) and ending with the highest upregulation (the darkest red color).

SPU ID	Gene name	Gene description	log2FC	Spatial expression	Reference
SPU_018393	Sp-Scratch	scratch homolog zinc finger protein	-2.604	L. variegatus: ciliary band	Slota et al., 2019
SPU_007147	Sp-Ngn	neurogenin	-2.291	apical ectoderm, ciliary band	Perillo et al., 2018
SPU_011755	Sp-Wnt9	wingless-type MMTV integration site family member 9	-2.029	unknown	n.a.
SPU_010167	Sp-Thytrprhr	Thyrotropin-releasing Hormone Receptor	-1.844	Gastrula: ciliary band	Arnone, unpublished; Wood et al., 2018
SPU_012548	Sp-Sparc_2	osteonectin, SPARC	-1.810	unknown	Livingston et al., 2006
SPU_025632	Sp-Pou4f2	Sp-Brn3, POU class 4 homeobox	-1.676	unknown	n.a.
SPU_012699	Sp-Nkx6-1	NK6 homeobox 1	-1.623	Gastrula: hindgut, prysm: hindgut and future pyloric sphincter	Israel et al., 2016; Arnone, unpublished

SPU_028169	Sp-RasO	Ras family orphan	-1.586	L. variegatus - blastula: ectoderm and vegetal plate; gastrula: one or both lateral sides of ciliary band	Slota et al., 2019
SPU_004599	Sp-Pitx2	paired-like homeodomain 2	-1.514	P. lividus: gastrula: ectoderm on the right side and near the tip of the archenteron	Duboc et al., 2005
SPU_017019	Sp-Risl2	Ras-like family 12 Like 2	-1.485	unknown	n.a.
SPU_002328	Sp-Nos_1	nitric oxide synthase	-1.477	unknown	n.a.
SPU_000276	Sp-PaxC	paired box C	-1.454	unknown	n.a.
SPU_016786	Sp-Isx	Sp-Rhox3, intestine specific homeobox-like	-1.429	Gastrula: hindgut	Arnone, unpublished
SPU_012050	Sp-Z11	zinc finger protein 11	-1.407	unknown	Materna et al., 2006
SPU_005955	Sp-Notchl1	Notch-like, fibropellin-like, Sp-Notch2-9	-1.361	unknown	n.a.
SPU_022846	Sp-FoxABL	forkhead box A and B-like, SpFoxAB-like	-1.357	Gastrula: oral ectoderm	Arnone, unpublished, Tu et al., 2006
SPU_024715	Sp-Cdx	caudal type homeobox	-1.344	Gastrula: hindgut	Arnone et al., 2006; Annunziata et al., 2013
SPU_003649	Sp-Z103	zinc finger protein 103, KLF-like, Kruppel-like factor-like	-1.325	unknown	Materna et al., 2006
SPU_002816	Sp-Hb9	homeobox HB9, HlxB9, Mnx1, motor neuron and pancreas homeobox 1	-1.323	P. lividus - gastrula: hindgut	Di Bernardo et al., 2000
SPU_028905	Sp-Pger2ep4	PGER2EP4	-1.255	unknown	n.a.
SPU_016685	Sp-Ap2	transcription factor AP-2, Sp-Tfap2	-1.186	L. variegatus prysm: oral ectoderm near ciliary band	Slota et al., 2019
SPU_019366	Sp-BsxL	brain-specific homeobox, bsh homolog, Sp-BsxL	-1.181	unknown	n.a.

SPU_020803	Sp-Trk	neurotrophic tyrosine receptor kinase, Ntrk, Ntrk1/2/3-like, TrkA/B/C-like	-1.151	L. variegatus - prysm: postoral neuron	Slota et al., 2019
SPU_020472	Sp-Hypp_975	hypothetical protein-975	-1.149	unknown	n.a.
SPU_023730	Sp-Isl	insulin gene enhancer protein	-1.119	Glastula and pluteus: CB, foregut, anus	Perillo et al., 2018; Annunziata et al., 2013
SPU_003175	Sp-Spec2d	Spec2d	-1.117	Gastrula: most likely aboral ectoderm	Chai et al., 1994
SPU_014418	Sp-FoxD	forkhead box D, Sp-FoxD_1	-1.104	Gastrula: apical ectoderm, oral side of the hindgut, oral side of the foregut	Tu et al., 2006
SPU_015587	Sp-Notum	Palmitoleoyl-Protein Carboxylesterase homolog (Drosophila)-1, Wingful, WF	-1.090	unknown	n.a.
SPU_001628	Sp-Wnt3	wingless-type MMTV integration site family, member 3	-1.081	L. variegatus - gastrula: hindgut and faint foregut signal	McClay et al., 2018
SPU_007599	Sp-Glass2	glass-like, Sp-Glass2	-1.074	unknown	n.a.
SPU_000749	Sp-Gcnf1	germ cell nuclear factor, Gcnf, nuclear receptor subfamily 6 group A member 1, Nr6a1	-1.063	unknown	n.a.
SPU_016443	Sp-Brn1-2-4	POU class 3 homeobox, Pou3f, Sp-Brn1-2-4, Sp-brn124	-0.969	Gastrula and pluteus: CB, neurons, foregut, stomodeum	Perillo et al., 2018; Cole and Arnone 2009
SPU_023894	Sp-Foxl	forkhead box l	-0.957	Gastrula: hindgut	Tu et al., 2006
SPU_014607	Sp-Spec2ce1-3	Spec2c	-0.930	Gastrula: most likely aboral ectoderm	Chai et al., 1994
SPU_008353	Sp-Sim	single-minded (sim) homolog (Drosophila), Sim1/2	-0.928	unknown	n.a.
SPU_016881	Sp-SoxE	SRY (sex determining region Y)-box E, Sox8/9/10	-0.897	Gastrula: veg2 derived mesoderm (SMC)	Juliano et al., 2006; Luo and Su, 2012, Annunziata et al., 2013

SPU_026687	Sp-Spec2a	spec2a	-0.897	Gastrula: aboral ectoderm	Chai et al., 1994
SPU_018136	Sp-Notch1-14	Notch 1 homolog 14	-0.883	unknown	n.a.
SPU_005955	Sp-Notchl1	Notch ligand-like 1, Notch-ligand1, Sp-Notchl1	-0.832	unknown	n.a.
SPU_011576	Sp-Nr1m2	nuclear receptor subfamily 1 group M member 2, NR1-like	-0.818	unknown	n.a.
SPU_024877	Sp-Osr	odd-skipped-related, z121	-0.814	Mid-gastrula: midgut-hindgut boundry and hindgut	Materna et al., 2006
SPU_006637	Sp-Z193	zinc finger protein 193	-0.798	unknown	Materna et al., 2006
SPU_027773	Sp-Unc44_262	ankyrin2,3/unc44-262	-0.796	unknown	n.a.
SPU_024307	Sp-NfIL3	nuclear factor, interleukin 3 regulated-like	-0.763	unknown	Howard-Ashby et al., 2006
SPU_023368	Sp-Asb5	ankyrin repeat and SOCS box-containing 5	-0.733	Early/mid gastrula: developing gut	Zazueta-Novoa & Wessel, 2014
SPU_013304	Sp-Unc44_229	ankyrin2,3/unc44-229	-0.730	unknown	n.a.
SPU_016449	Sp-Hnf6	onecut homeobox, Onecut1/2/3-like	-0.714	Gastrula: ciliary band	Otim et al., 2004
SPU_015772	Sp-Klf15	Kruppel-like factor 15-like, z174	-0.695	unknown	Materna et al., 2006
SPU_002815	Sp-Dlx	distal-less homeobox, Dll	-0.691	Blastula: veg1 ecto, aboral ecto; Gastrula: possible weak detection	Howard-Ashby et al., 2006
SPU_023727	Sp-Klf13	ruppel-like factor 13-like, z188, Klf9/13/14/16-like	-0.657	Blastula: apical ectoderm and vegetal plate; gastrula: apical and oral ectoderm	Materna et al., 2006
SPU_007452	Sp-Spec1	Spec 1a	-0.653	Blastula: aboral ectoderm	Hardin et al., 1985
SPU_022757	SPU_022757	unknown	-0.643	unknown	Tu et al., 2012
SPU_022683	Sp-Drd1_2	Dopamine Receptor D1	-0.611	unknown	Tu et al., 2012

SPU_013704	Sp-Unc4.1	UNC homeobox, Uncx, Uncx4.1, Sp-Unc4.1_1	-0.587	Blastula: vegetal plate, apical plate; Gastrula: foregut	Howard-Ashby et al., 2006
SPU_016016	Sp-NotchI5_1	Notch ligand-like 5, Sp-NotchI5_1	0.597	unknown	n.a.
SPU_002783	Sp-NotchL_1	notch-like-1	0.600	unknown	n.a.
SPU_006803	Sp-Creb3l3	cAMP responsive element binding protein 3-like, Sp-Creb3	0.636	unknown	Howard-Ashby et al., 2006
SPU_006917	Sp-Runt1	runt-related transcription factor-like-1	0.661	Blastula: vegetal plate, PMC, aboral ecto; Glastula: gut, oral ectoderm, SMC	McCarty & Coffman, 2013; Robertson et al., 2008
SPU_019002	Sp-FoxQ2_1	forkhead box Q2 (copy 2)	0.682	Blastula and gastrula: apical ectoderm	Yaguchi et al., 2008; Tu et al., 2006; Range et al., 2017
SPU_020311	Sp-fi	Kruppel-like factor 2/4-like, z85	0.688	Blastula: aboral ectoderm or oral ectoderm, mid-gastrula: aboral ectoderm and/or oral ectoderm	Materna et al., 2006
SPU_014628	Sp-Erf	Ets2 repressor factor-like, ERF-like, Etv3-like, Mets-like	0.714	Ubiqutous throughout development	Rizzo et al., 2006
SPU_025302	Sp-Alx1	aristaless-like homeobox 1-like	0.716	Blastula and gastula: PMC	Ettenson et al., 2003
SPU_008975	SPU_008975	unknown	0.726	unknown	Tu et al., 2012
SPU_018483	Sp-Erg	v-ets oncogene related (Erg)-like, Friend leukemia integration 1 (Fli1)-like	0.746	Blastula: PMC and SMC; gastrula: SMC	Rizzo et al., 2006
SPU_011004	Sp-Unc44L_68	ankyrin2,3/unc44-like-68	0.761	unknown	Tu et al., 2012
SPU_028093	Sp-Scl	stem cell protein-like	0.805	Blastula: veg2 meso	Solek et al., 2013
SPU_017837	Sp-Nkx3.2	Sp-Homeo4, homeodomain containing 4	0.821	Gastrula: oral animal pole and foregut	Wei et al., 2011
SPU_003105	Sp-Hypp_1433	hypothetical protein-1433; G protein-coupled receptor	0.827	unknown	Tu et al., 2012

SPU_012448	Sp-Ovo	ovo-like, z157	0.841	Blastula: PMC; Glastula: no detection	S.C. Materna et al., 2006; C.E Juliano et al., 2006
SPU_011246	Sp-IrxB	iroquois homeobox B	0.853	unknown	Annunziata & Arnone 2014
SPU_004526	Sp-Nr1h6c	nuclear receptor	0.858	unknown	Kontrogianni-Konstantopoulos et al., 1996
SPU_020411	SPU_020411	unknown	0.887	unknown	Tu et al., 2012
SPU_007638	Sp-Ikke	IKK epsilon, inhibitor of nuclear factor kappa B kinase epsilon	0.912	unknown	Tu et al., 2012
SPU_010776	Sp-Unc5	Netrin receptor UNC5	0.947	unknown	Whittaker et al., 2006; Tu et al., 2012
SPU_002874	Sp-Ets1/2	v-ets erythroblastosis virus E26 oncogene homolog 1/2-like	0.959	Blastula: veg2 meso, PMC; Glastula: SMC, PMC	Rizzo et al., 2006
SPU_026877	Sp-Irf4	interferon regulatory factor 4-like	0.990	unknown	Cui et al., 2014; Peter & Davidson, 2009; Hibino et al., 2006; Howard-Ashby et al., 2006
SPU_002129	Sp-Not	notochord homeobox-like, Noto, flh	1.031	Blastula: SMC, oral ectoderm; Gastrula: tip of the archenteron	Howard-Ashby et al., 2006; Barsi et al., 2015
SPU_028061	Sp-Dach	dachshund (dac) homolog, Sp-Dach1, Sp-Dac	1.032	Blastula: endoderm; Gastrula: hindgut, midgut, foregut	Howard-Ashby et al., 2006
SPU_011396	Sp-Znf294	zinc finger protein 294	1.047	unknown	Tu et al., 2012
SPU_012384	Sp-FoxQ2	forkhead box Q2	1.078	Blastula and gastrula: apical ectoderm	Yaguchi et al., 2008; Tu et al., 2006; Range et al., 2017
SPU_012240	Sp-Znf259l	zinc finger protein 259	1.118	unknown	Tu et al., 2012
SPU_003102	Sp-Jun	jun proto-oncogene-like, AP-1-like	1.153	Blastula: PMC; Glastula: PMC	Russo et al., 2014; Dylus et al., 2016; Russo et al., 2018

SPU_021119	Sp-MyoD	myogenic differentiation 1-like	1.187	Blastula:and Glastula: PMC	Beach et al., 1999; Andrikou et al., 2013
SPU_026657	Sp-Unc44_350	ankyrin2,3/unc44-350	1.188	unknown	Tu et al., 2012
SPU_027734	Sp-Cbfb	SpCBFbeta, Core Binding Factor Beta, PEBP2beta	1.224	Ubiqutous throughout development; gastrula: enriched in the hindgut and oral ectoderm	Robertson et al., 2006
SPU_006032	Sp-Hypp_1626	hypothetical protein-1626	1.273	unknown	Tu et al., 2012
SPU_022573	Sp-Nk7	NK7 homeobox	1.353	Blastula and gastrula: PMC	Dylus et al., 2016; Rafiq et al., 2014
SPU_020451	Sp-Bra	Brachyury	1.414	Blastula: veg endo, oral ectoderm; Gastula: blastopore, hindgut, stomodeum	Gross & McClay, 2001; Croce et al., 2001; Rast et al., 2002
SPU_010404	Sp-Irf	interferon regulatory factor, Sp-irf1	1.431	unknown	Cui et al., 2014; Peter & Davidson, 2009; Hibino et al., 2006; Howard-Ashby et al., 2006
SPU_010403	Sp-FoxY	forkhead box Y, forkhead C-like	1.597	Blastula: small micromeres, SMC; gastrula: SMC, tip of the archenteron	Materna et al., 2013; Annunziata & Arnone, 2014; Andrikou et al., 2015
SPU_009427	Sp-Z225	zinc finger protein 225	1.862	unknown	Materna et al., 2006
SPU_025584	Sp-Tbr	T-box brain-like, Eomes-like, Tbx21-like, ske-T	1.967	Blastula and gastrula: PMC	Sharma & Ettensohn, 2010
SPU_027025	Sp-Znfx1-17	zinc finger, NFX1-type containing 1-17	2.144	unknown	Tu et al., 2012
SPU_024900	Sp-Drd2	dopamine receptor D2	2.228	unknown	Tu et al., 2012

SPU_021947	Sp-Krl	zinc finger protein 474, Sp-Z13, transcriptional repressor Krl (SpKrl)-like	2.573	Blastula: endoderm; Gastrula: not detected	Stepicheva et al., 2015; Materna et al., 2006; Minokawa et al., 2004; Howard et al., 2001
SPU_026196	SPU_026196	unknown	2.581	unknown	Rafiq et al., 2014; Tu et al., 2012
SPU_017649	Sp-Hgf	hepatocyte growth factor-like, plasminogen-like	2.973	unknown	Lapraz et al., 2006; Tu et al., 2012

The absence of Brachyury changed the normal pattern of expression of 23 ectodermal genes, from which 8 are expressed in the oral ectoderm, 6 in the aboral, 5 in the apical ectoderm, and 6 in the ciliary band region. 12 mesodermal genes were under the influence of Brachyury knock-down, from which 6 are known to be normally expressed in SMCs, and 7 in the PMCs. Moreover, 3 ubiquitously expressed genes were upregulated, while 14 endodermal genes were both up- and down-regulated.

The effect of the Brachyury knock-down after 48hpf seems to be much larger at the end than on the onset of gastrulation, judging from the number of affected genes. These results show that Brachyury is involved in regulating all three germ layers throughout the whole process of gastrulation.

However, in order to find the putative direct and indirect Brachyury targets, the following sections of this Chapter will be focused on describing affected TFs and signaling molecules which expression patterns are known from the published literature, unpublished in situ hybridization experiments, targets found by the unpublished Brachyury 24hpf and 48hpf ChIP-seq datasets from Arnone's lab. The final proposed list of the putative direct and indirect Brachyury targets contains the integrated data collected from the different sources mentioned above.

4.2.5 Combing differentially expressed genes after Brachyury knock-down with Brachyury ChIP-seq and wild type ATAC-seq datasets

Since the differential transcriptomic analysis after the Brachyury knock-down does not give any information on whether the effect of Brachyury on the downstream transcription factors and signaling molecules is direct or indirect, they were combined with ChIP-seq/ATAC-seq datasets at both developmental stages. The intersected peaks were used to search for the presence of peaks close to the genes discovered by the differential transcriptomic approach described in the previous paragraphs in order to find out which of the putative targets are under the direct influence of Brachyury.

De novo *motif discovery from Brachyury ChIP-seq datasets*

To detect which peaks contain Brachyury binding sites, ChIP-seq/ATAC-seq intersected bed files were scanned for Brachyury motifs. To this aim, de novo motif discovery was performed with the higher stringency 24hpf ChIP-seq peak file, which was not intersected with the ATAC-seq peak file.

To find short recurring motifs with fixed-length patterns, this task was accomplished using the *DREME* tool (Bailey, 2011). The motif discovered is represented in Figure 4.5.

Figure 4.5 *De novo* motif discovered from 24hpf *Strongylocentrotus purpuratus* Brachyury ChIP-seq dataset with *DREME*. (Bailey, 2011)

The obtained motif had a length of 6 nucleotides. Its matrix was searched for similarity matching using the Jaspar V3 database (Bryne *et al.*, 2008) and TRANSFAC database (Matys *et al.*, 2003) with the *STAMP* tool (Mahony and Benos, 2007; Mahony, Auron and Benos, 2007) as described in Chapter 2. The first closest match coming from the Jaspar database was a T (Brachyury) motif, aligned to the obtained motif with an e-value equal to 1.8814e-07 (Figure 4.6), and the first two closest matches found in the TRANSFAC database were vertebrate Tbx5, aligned with the e-value equal to 5.0622e-08 and Brachyury motif (Figure 4B), aligned with the e-value equal to 2.9866e-06 (Figure 4.6).

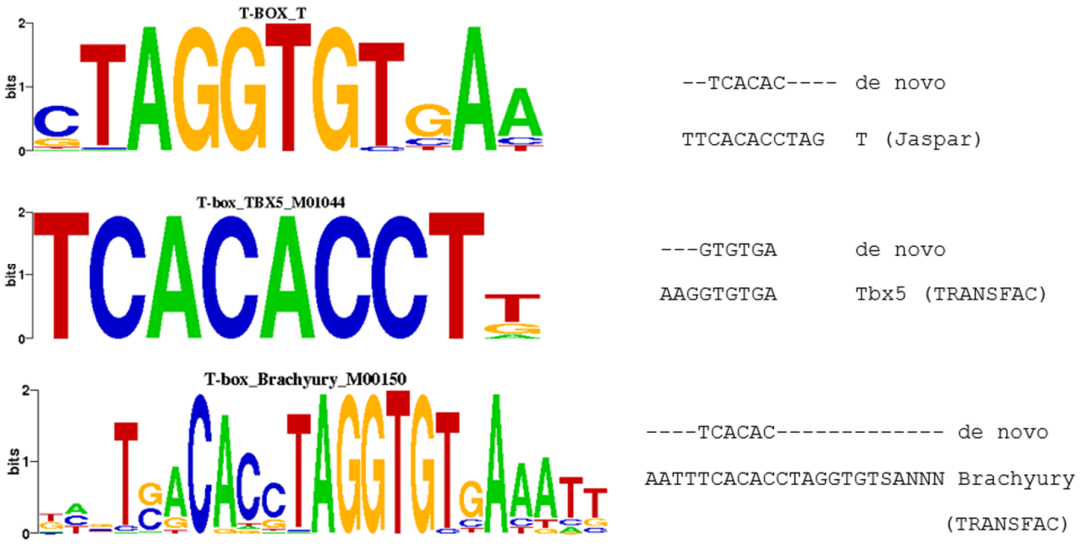


Figure 4.6 *De novo* discovered Brachyury motif from the *S. purpuratus* 24hpf ChIP-seq experiment aligned to Jaspar database and TRANSFAC database using *STAMP* tool. The closest matches after the alignment to Jaspar database was T (Brachyury) and the closest matches after the alignment to TRANSFAC database were Tbx5 and Brachyury.

The reason behind choosing *de novo* discovery of enriched short motifs in large omics datasets lies in the observation that many transcription factors can bind their specific motifs in "half-form." In the case of Brachyury, it does not need to bind to a palindromic sequence, contrary to what was first discovered *in vitro* using

systematic evolution of ligands by exponential enrichment (SELEX) method (Kispert and Herrmann, 1993). *In vitro* conditions allow TFs to interact with duplicated (or palindromic) nucleotides more efficiently, and probably, the motif discovery could be biased. Since Chromatin Immunoprecipitation allows detection of TF bound to the chromatin *in vivo*, more accurate results are expected.

In conclusion, the alignments of *de novo* discovered *S. purpuratus* Brachyury motif point to its conservation with vertebrate motifs since the discovered motif contains the core T-box domain, which seems to be consistent with the known vertebrate Brachyury motif, that was discussed in Chapter 1 (Kispert, Koschorz and Herrmann, 1995; Sebé-Pedrós and Ruiz-Trillo, 2017) and (Sebé-Pedrós, Ariza-Cosano, *et al.*, 2013).

Moreover, a high number of genes are affected by Brachyury perturbation, which could be explained by a complex cascade of events associated with directly affected genes and the effect on their downstream targets.

Intersection of Brachyury ChIP-seq datasets with ATAC-seq datasets and proposed direct targets

The unfiltered 24hpf ChIP-seq dataset contained 15274 peaks, while intersected ChIP-seq/ATAC-seq at the same time point file had lowered the number of peaks down to 1069. The unfiltered 48hpf ChIP-seq dataset contained 11230 peaks, and after the intersection with the 48hpf ATAC-seq datasets, the number of peaks was lowered to 511. The peaks were annotated with names of the genes found proximal to these peaks using the *S. purpuratus* genome v3.1 and its gene annotation (Kudtarkar and Cameron, 2017).

Differentially expressed genes discovered at 27hpf after perturbation of Brachyury were searched for in the intersected ChIP/ATAC-seq dataset corresponding to 24hpf. From the 279 differentially expressed genes at 27hpf, 36 contained intersected peaks adjacent to them. 15 out of 36 peaks were found to have a Brachyury motif. (For more information, refer to the Non-book component of this thesis). 18 out of 36 differentially expressed TFs/signaling molecules were found to contain at least one peak, and 9 out of 18 of them contained at least one peak with

a Brachyury motif, while 9 genes contained peaks without a Brachyury motif. The annotated peaks were proposed to be direct Brachyury targets, and they are represented in Table 4.3. For more details, refer to the non-book component files.

Table 4.3. Putative Brachyury direct targets at early gastrula stage of *S. purpuratus*. Each gene represented was found to contain at least one peak in ChIP-seq data set and a possible function as a transcriptional activator or repressor is proposed.

Gene	ChIP whole embryo/Brachyury motif	Number of peaks	Brachyury function proposed by DEG analysis
Sp-Isl	yes/no	4	activator
Sp-Nk1	yes/yes	2	activator
Sp-Hox7	yes/no	1	activator
Sp-Spec1	yes/yes	2	activator
Sp-IrxA	yes/no	2	activator
Sp-FoxA	yes/yes	2	activator
Sp-Spry/Socscp	yes/yes	1	repressor
Sp-Spalt	yes/no	2	repressor
Sp-Ese	yes/yes	1	repressor
Sp-Runt1	yes/no	2	repressor
Sp-Ovo	yes/no	1	repressor
Sp-Delta	yes/no	3	repressor
Sp-Scl	yes/no	1	repressor
Sp-Bra	yes/yes	6	repressor
Sp-Fra2	yes/yes	1	repressor
Sp-GataC	yes/no	1	repressor
Sp-Egr	yes/yes	1	repressor
Sp-Jun	yes/no	1	repressor

Differentially expressed genes discovered at 48hpf after Brachyury perturbation were searched for in the intersected ChIP/ATAC-seq 48hpf dataset. Out of 1088 differentially expressed genes, 36 were found to contain peaks in ChIP/ATAC-seq, from which 9 had Brachyury motif sequence. (For more details, refer to the non-book component files)

Considering the 98 detected TFs/signaling molecules, 13 had peaks in the ChIP/ATAC-seq dataset, from which 3 peaks contained Brachyury motif. Those 13 TFs/signaling molecules are likely to be direct Brachyuy targets, and they are represented in Table 4.4.

Table 4.4 Putative Brachyury direct targets at late gastrula stage of *S. purpuratus*. Each gene represented was found to contain at least one peak in ChIP-seq data set and a possible function as a transcriptional activator or repressor is proposed.

Gene	ChIP whole embryo/Brachyury motif	Number of peaks	Brachyury function proposed by DEG analysis
Nkx6-1	yes/no	1	activator
Cdx	yes/no	1	activator
Ap2	yes/yes	2	activator
Isl	yes/no	1	activator
FoxD	yes/no	1	activator
SoxE	yes/no	2	activator
Osr	yes/no	1	activator
Hnf6	yes/no	1	activator
Drd1_2	yes/no	1	activator
Scl	yes/no	1	repressor
Unc5	yes/yes	2	repressor
Not	yes/no	2	repressor
Dach	yes/no	1	repressor

It could be speculated that in late gastrula, Brachyury might have a preference in binding to secondary motif sequences due to the lower number of Brachyury motif sequences found in the 48hpf ChIP/ATAC-seq peaks, compared to the 27hpf ChIP/ATAC-seq dataset. It has been shown that it is not rare to find that a TF can recognize and bind to multiple motifs. The appearance of secondary motifs can happen if the TF interacts with a non-DNA binding cofactor (Siggers *et al.*, 2011) or with another TF (Joshi, Sun and Mann, 2010; Slattery *et al.*, 2011). For instance, the interaction between homeobox TFs Exd and Hth results in this complex recognizing different motifs compared to their individual binding motifs (Slattery *et al.*, 2011).

4.2.6 Analysis of expression patterns of the putative Brachyury targets based on the integrative approach

In order to reconstruct the GRN around Brachyury based on RNA-seq and ChIP-seq data, differentially expressed transcription factors and signaling molecules detected after the injection of Brachyury morpholino were searched for in the literature. Moreover, to this aim, *in situ* hybridization expression patterns available were studied to gain a clear idea about the spatial expression of the proposed targets. Here, I present the potential candidates for the GRN.

Brachyury targets at 27hpf

Genes under putative control of Brachyury in the ectoderm. There are 7 genes proposed to be targets of Brachyury in the oral ectoderm. Based on the observed upregulation from the differential RNA-seq data, it seems that Brachyury a gene with a self-regulatory feedback loop. This is supported by the ChIP/ATAC-seq data, which shows that there are 6 peaks around Brachyury at 27hpf, 2 in the promoter, and 4 in the gene body region. Moreover, the *in situ* hybridization after the Brachyury morpholino also supports this hypothesis. Figure 4.8 shows that Brachyury is upregulated in both domains – oral ectoderm and the vegetal plate.

FoxA belongs to the Forkhead family. It was isolated by Harada *et al.* from the sea urchin *H. pulcherrimus,* while from *S. purpuratus,* it was isolated by Tu *et al.* in 2006. Its expression was confirmed to be present in oral ectoderm and veg2 endoderm lineage before gastrulation, while in gastrulae, its expression domain is spread throughout the gut (Harada *et al.*, 1996; Tu *et al.*, 2006). *FoxA* is a well-known Brachyury target (Oliveri *et al.*, 2006; de-Leon and Davidson, 2010). It appears to be down-regulated after the Brachyury MASO, and there are 2 ChIP/ATAC-seq peaks found in its proximity, one in the promoter that contains Brachyury motif and the second one in its only exon. In addition, *FoxA* and *brachyury* were assessed by *in situ* hybridization in normal and perturbed conditions. In normal conditions, at 24hpf, *brachyury* and *FoxA* are entirely overlapping in their expression in the veg2 endoderm cells. (Figure 4.7). At 27hpf, *FoxA* is co-expressing with *brachyury* in both domains – the oral ectoderm (Figure 4.8A) and the vegetal plate (Figure 4.8C). When Brachyury protein is knocked-down, it is evident that *FoxA* expression is

almost completely abolished in the oral ectoderm (Figure 4.8A'), which has been described before in *P. lividus* (Saudemont *et al.*, 2010). The expression is still present in the vegetal plate, but at a much lower level (Figure 4.8 C'). On the other hand, *Brachyury* shows high upregulation in both oral ectoderm (Figure 17 A') and vegetal plate (Figure 4.8 C').

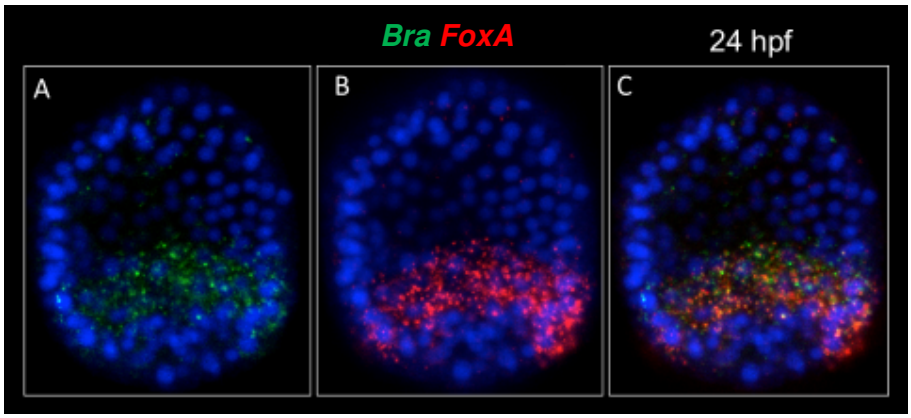


Figure 4.7. Double whole-mount fluorescent *in situ* hybridization of *Bra* and *FoxA* at 24hpf in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and it completely co-expresses with *FoxA* which is represented in red, in the vegetal plate of the embryo. A) *brachyury*; B) *FoxA*; C) merged. Courtesy of Periklis Paganos.

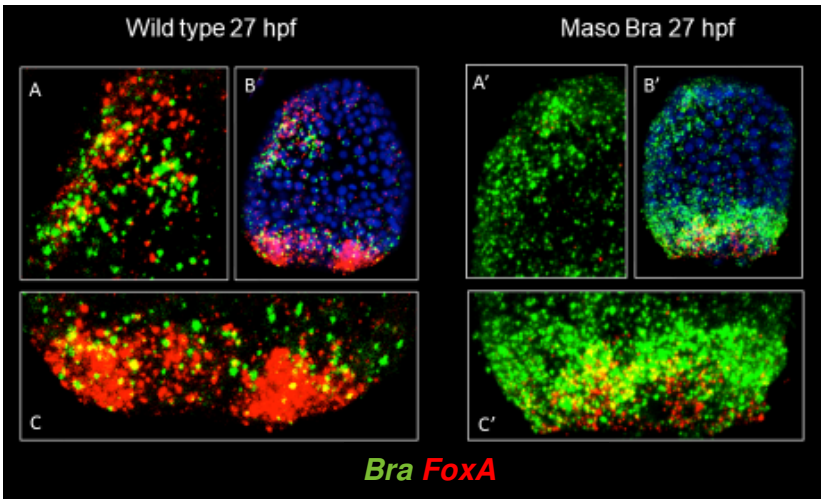


Figure 4.8. Double whole-mount fluorescent *in situ* hybridization of *FoxA* and *Bra* at 27hpf in normal and Brachyury perturbed condition in *S. purpuratus*. *Brachyury (Bra)* is represented in green, and *FoxA (FoxA)* is represented in red. A), B) and C) wild type condition; A'), B'), and C') Brachyury knock-down embryos. A) and A') oral ectoderm, B) and B') whole embryo; C) and C') vegetal plate. *FoxA* and *brachyury* co-express in oral ectoderm and vegetal plate. Courtesy of Periklis Paganos.

Nk1 transcripts are accumulated at the mesenchyme blastula stage in the oral ectoderm veg1 lineage at the border with veg1 endodermal lineage (Minokawa *et al.*, 2004; Su *et al.*, 2009). During the gastrula stage, the expression stays in this domain and also appears in the veg1 endoderm located on the oral side, represented in Figure 4.12 (Minokawa *et al.*, 2004; Su *et al.*, 2009). *Nk1* was already proposed as a direct Brachyury target by differential macroarray screening in the oral ectoderm by Rast *et al.*, 2002. 24hpf ChIP-seq dataset contains two peaks around *Nk1*, one in the intronic and one in the intergenic region upstream of the TSS. The intergenic region is most likely the regulatory region where the Brachyury transcription factor binds since the Brachyury motif is detected in this peak. This places *Nk1* as a direct target gene of Brachyury.

Runt1 was upregulated and contained ChIP-seq peaks in two intronic positions (Coffman et al., 1996; Robertson et al., 2002). *Runt1* expression pattern is ubiquitous during the embryonic stages (Robertson et al., 2002). Bra may control its levels in both endoderm and ectoderm by repressing it (Table 4.1 and 4.3).

Elk belongs to the Ets family of transcription factors, and it was first isolated by Rizzo *et al.* (Rizzo *et al.*, 2006) Its expression pattern seems to be ubiquitous during development; however, at 24hpf, it looks like that the amounts of transcripts are enriched in the vegetal plate and absent from the ectoderm (Rizzo, Coffman and Arnone, 2016). Since it was detected as upregulated in the 27hpf Bra MASO RNA-seq dataset, it could be suggested that Brachyury regulates its expression in the oral ectoderm by repressing it.

IrxA, known to be expressed in the aboral ectoderm (Howard-Ashby, Stefan C Materna, *et al.*, 2006; Su *et al.*, 2009; Ben-Tabou de-Leon *et al.*, 2013), was found down-regulated after the Brachyury knock-down. ChIP/ATAC-seq data allows us to infer that it is a direct Bra target since there is one peak detected in the intergenic region downstream of *IrxA*. It is possible that at 27hpf, there is a positive input of Brachyury on *IrxA* in the oral ectoderm. This is supported by the observation in *P. lividus*, where it seems that *IrxA*, apart from being expressed in the aboral ectoderm, just at the beginning of gastrulation, starts to be expressed in the small region of oral ectoderm (Saudemont *et al.*, 2010).

Isl (*islet*) is a transcription factor expressed in "pancreatic-like" cells (Perillo *et al.*, 2018). In this work, *Isl* expression was tested at the earlier stage of development, 24hpf, and later, at 48hpf. In the blastula stage, at 24hpf, it seems that *Isl* is co-expressed with *brachyury* in both domains, the oral ectoderm (shown in Figure 4.9A, B, C) and vegetal plate (shown in Figure 4.9D, E, F). ChIP/ATAC-seq dataset contains four peaks close to the *Isl*, two upstream and two peaks downstream of the TSS. Therefore, Isl is likely to be a direct Brachyury target.

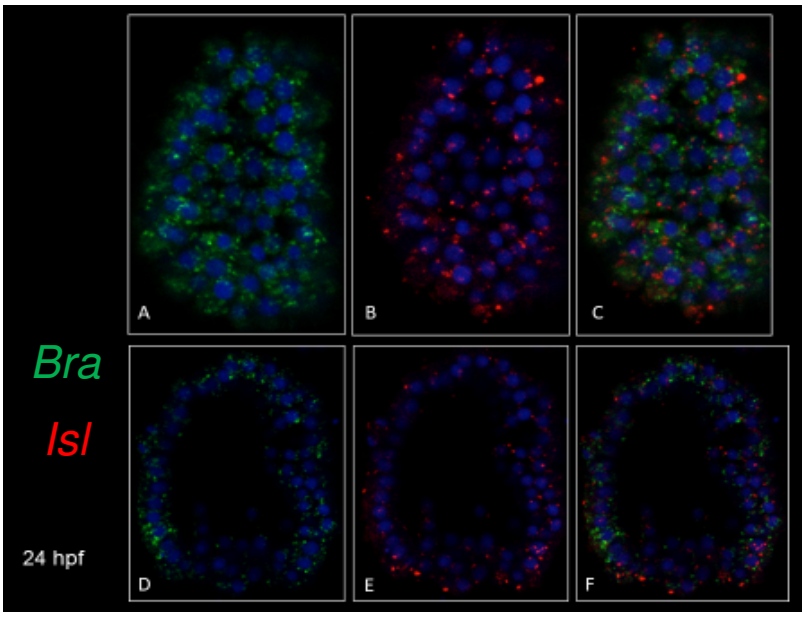


Figure 4.9 Double whole-mount fluorescent *in situ* hybridization of *brachyury* and *Isl* at 24hpf in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and co-expresses with *Isl,* which is represented in red, in the region of oral ectoderm (C), and vegetal plate (C and F). A) and D) *Brachyury*; B) and E) *Isl*; C) and F) merged. Courtesy of Periklis Paganos.

Multiple *Spec* genes were affected after Brachyury knock-down, as shown in Table 4.1 and Table 4.2. They are known as the major calcium-binding proteins, from which *Spec1* and *Spec2d* were tested for spatial expression, showing that they are aboral ectoderm specific (Hardin *et al.*, 1985). All *Spec* genes affected by Brachyury perturbation showed down-regulation. In addition, the 24hpf ChIP-seq data set shows the existence of two peaks containing Brachyury binding sites, both located in the intergenic region upstream of the *Spec1* TSS. In the 48hpf ChIP-seq dataset, no peaks around *Spec1* were found. These results seem to be contradictory.

However, it could be possible that in the blastula *Spec1* gene is expressed at lower levels in the oral ectoderm, where Brachyury could be activating it, but due to the low expression, it could not be detected by *in situ* hybridization. Later downregulation of *Spec1* in the late gastrula, together with *Spec2d*, *Spec2ce1*-3, *Spec2a* could be explained by the "domino-effect" of multiple ectodermal genes affected directly and indirectly by the absence of Brachyury transcription factor.

Hox7 is an aboral ectoderm marker that was found to contain one ChIP-seq peak in the intergenic region upstream of its TSS.

In the case of *Spec1* and *Hox7*, which are aboral ectoderm markers, it seems contradictory to find that Brachyury is a potential activator and that there were ChIP-seq peaks detected around them. Different hypotheses could explain this:

1. It could be possible that Brachyury is activating those genes, but again the sensitivity of the *in situ* hybridization was unable to detect the presence of their transcripts in the oral ectoderm.
2. It could be possible that Brachyury is present at the border regions between the oral and the aboral ectoderm that as well could not be detected by *in situ* hybridization.
3. It could be possible that the peaks around these genes detected by the ChIP-seq were false positives.
4. It could be possible that Brachyury indeed binds close to the mentioned genes, but it is not involved in their activation, but it could be involved in the regulation of the unknown distally located genes whose regulatory regions are located close to *Hox7* and *Spec1*.

The only way to answer these questions is to perform the actual co-expression and functional experiments. Moreover, taking into account the phenotypic differences in the overall embryo morphology at both 27 and 48hpf, it is clearly visible that ectoderm shows malformations. The consequence of the Brachyury absence could be explained by the small disruptive effect in the oral-aboral axis formation, which could also influence the down-regulation of the mentioned aboral ectoderm genes. Therefore, these genes are not considered at the moment as direct Brachyury

targets. They could be, however, considered as indirect targets through unknown signaling events, together with the aboral ectoderm specific homeobox-NKL transcription factor *Dlx* (Howard-Ashby, Stefan C Materna, *et al.*, 2006) that was detected as downregulated at 48hpf, and apical ectoderm specific genes that showed a change in their expression after the absence of Brachyury, a zinc-finger transcription factor, *z133* (Materna *et al.*, 2006), and a signaling molecule *Sfrp1*/5 (Range and Wei, 2017) found in the apical plate.

At this moment, differentially expressed apical (*z133* and *Sfrp1*/5) and aboral ectoderm (*Hox7* and *Spec1*) genes are considered to be indirect targets of Brachyury through the still unknown signaling events (Hardin *et al.*, 1985; Howard-Ashby, Materna, *et al.*, 2006; Materna *et al.*, 2006; Range and Wei, 2017).

Genes under putative control of Brachyury in the endoderm. At 27hpf, there were 3 genes detected as being activated by Brachyury in the endoderm. Those are the already described *FoxA* and *Nk1*, which both contain two ChIP-seq peaks in their proximity with a Brachyury binding site. Considering the spatial and temporal pattern of *FoxA* and *brachyury* expression, *FoxA* is a direct target of Brachyury in the 21-24hpf veg2 GRN and 25-27hpf veg1 GRN. *Nk1* is considered a direct target in the regulatory state of veg2 endoderm 25-27hpf. *Apobec1* belongs to the cytidine deaminase enzyme family, which also has a function of DNA and RNA editing (Conticello *et al.*, 2005; Petit *et al.*, 2009). In sea urchins, its expression pattern is confined to the endodermal lineage, specifically to veg2 endo. This gene was down-regulated after the Bra perturbation, which confirms that it is Bra target in this cell lineage, as shown previously by Rast and colleagues (Rast *et al.*, 2002) but it does not have ChIP-peaks around it.

Genes under putative control of Brachyury in the mesoderm. Compared to only 3 endodermal TFs that were affected by the absence of Bra, 14 mesodermal TFs were affected. These TFs were all upregulated in the absence of Bra, and they are known to be either PMC and/or SMC specific genes. Secondary mesenchyme specific genes that were affected are *Prox1*, *Spalt*, *Delta*, *GataC*, *Ese*, *Slc*, *Elk*, and *Ets1*/2 (Materna *et al.*, 2006; Rizzo *et al.*, 2006; Materna and Davidson, 2012; Solek *et al.*, 2013). Primary mesenchyme genes that were affected are *Ets1*/2, *Alx1*, *MyoD1*, *Jun*,

Ovo, and *Fra2* (Ettensohn, 2003; Howard-Ashby, Stefan C. Materna, *et al.*, 2006; Materna *et al.*, 2006; Rizzo *et al.*, 2006; Andrikou *et al.*, 2013; Russo *et al.*, 2014). 8 of them contain ChIP-seq peaks around them, which can strongly suggest that Bra acts directly on them as a repressor (*Egr*, *GataC*, *Spalt*, *Ese*, *Scl*, *Jun*, *Fra2*, and *Ovo*). *Ese* is a transcription factor belonging to the Ets family, and it was described to be expressed in non-skeletogenic veg2 mesodermal cell lineage during development (Rizzo et al, 2006; Slota et al, 2019). The *in situ* hybridization experiment confirmed that there is no co-expression with *Brachyury* (Figure 4.10). There is one ChIP-seq peak detected in the promoter region of *Ese*, containing multiple Brachyury biding sites. Based on these observations, it can be inferred that Brachyury act as a repressor of Ese in veg2 endo lineage.

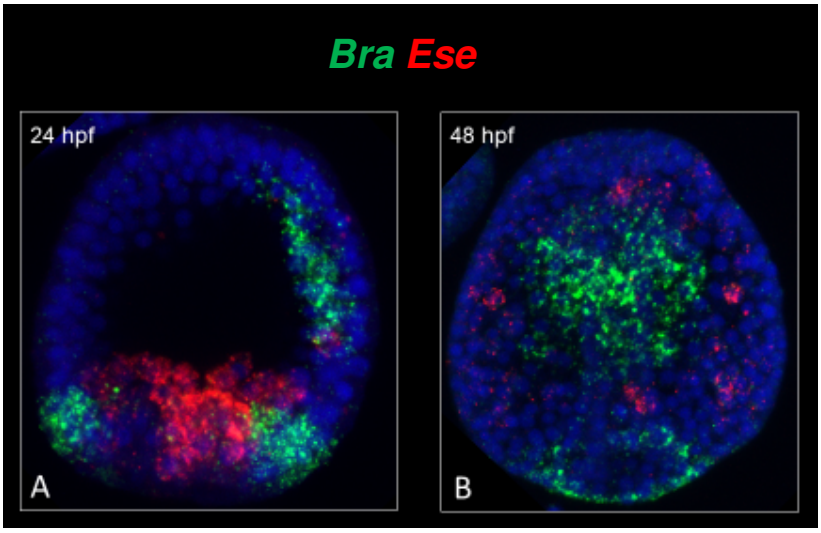


Figure 4.10 Double whole-mount fluorescent *in situ* hybridization of *Bra* and *Ese* at 24hpf and 48hpf in *S. purpuratus*. *Brachyury (Bra)* is represented in green, and *Ese* is represented in red. A) *Brachyury* domains of expression is in veg2 endodermal lineage and oral ectoderm, while *Ese* is expressed in veg2 meso lineage. B) *Brachyury* is expressed in the hindgut region and oral ectoderm, while *Ese* is expressed in non-skeletogenic mesoderm. Courtesy of Periklis Paganos.

Delta is one of the most conserved and most important signaling molecules among Metazoans, a part of a crucial signaling pathway involved in embryonic development, the Delta/Notch signaling pathway (Artavanis-Tsakonas, Rand and Lake, 1999). It has been shown that during sea urchin development *Delta*, *FoxA* and *Gcm* play a

fundamental role in the specification of endoderm and mesoderm. (Sweet, Gehring and Ettensohn, 2002; Croce and Mcclay, 2010; Peter and Davidson, 2011a). *Delta* is first expressed in the micromeres, and it starts to signal between the eighth and tenth cleavage to the veg2 cells (McClay et al., 2000; Sweet et al., 2002). Later on, in the unhatched blastula, the expression is restricted to veg2 endomesoderm cells, together with *Gcm* and *FoxA* (Croce and Mcclay, 2010). It has been proven that the continuous Delta signal is necessary prior to endoderm and mesoderm specification, precisely, just before hatching of the blastula (Croce and Mcclay, 2010). In the hatched blastula, *Delta* continues to be expressed in the future veg2 meso lineage (the inner ring of veg2 cells), promoting the expression of *Gcm*, before *Gcm* continues to maintain its own expression and to promote mesoderm specification (Croce and Mcclay, 2010). In this inner lineage, *FoxA* slowly starts to be down-regulated and it starts to be upregulated in the outer ring of veg2 lineage, where it will promote the endodermal fate (Oliveri *et al.*, 2006; Croce and Mcclay, 2010). In mesenchyme blastula, when the endodermal and mesodermal lineages are already separated, *Delta* is never co-expressing with *Brachyury* (Figure 4.11A).

In the gastrula stage, *Delta* expression is confined to the non-skeletogenic mesoderm (Figure 4.11B). In addition, the 24hpf ChIP-seq dataset shows three peaks in proximity to the *Delta* gene. Two are located in exon 1 and intron 10, and the third one is located in the intergenic region upstream of the *Delta* TSS. However, no Brachyury binding sites were detected in these three peaks. This could suggest that Brachyury acts with a cofactor and represses *Delta*, most probably in the veg2 endo lineage before 24hpf, when it starts to be expressed in the veg1 ring of cells.

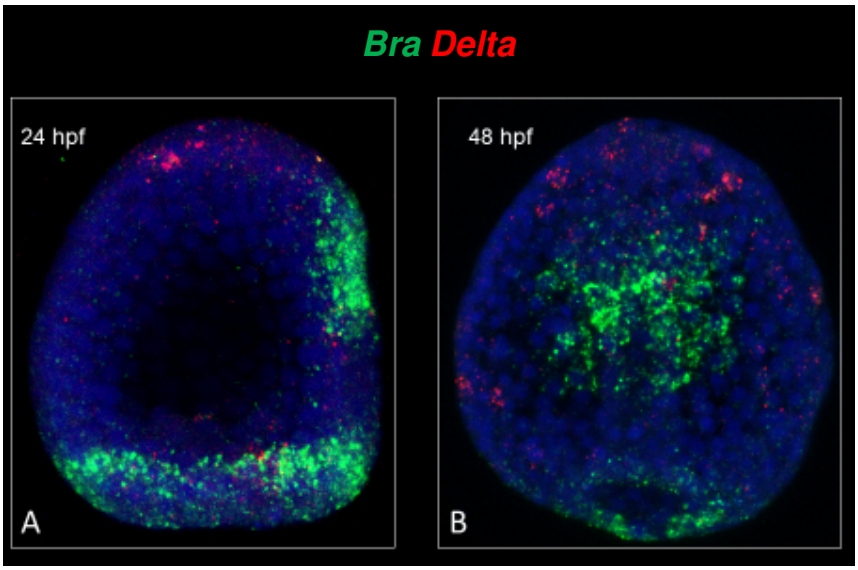


Figure 4.11 Double whole-mount fluorescent *in situ* hybridization of *Bra* and *Delta* at 24hpf and 48hpf in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and *Delta* is represented in red. A) *Brachyury* domains of expression are represented in veg2 endodermal lineage and oral ectoderm, while *Delta* is expressed in PMC and veg2 meso lineage. Ectodermal staining is most probably non-specific. B) *Brachyury* is expressed in the hindgut region and oral ectoderm, while *Delta* is expressed in non-skeletogenic mesoderm. Courtesy of Periklis Paganos.

Moreover, *Egr*, an ectodermal transcription factor (Materna *et al.*, 2006), seems to be repressed by Brachyury in endoderm, too.

These results suggest that, while Brachyury is expressed in the veg2 endodermal lineage from 21 to 24hpf, it will most probably repress the mesodermal genes. The known mesodermal fate repressor FoxA is likely to co-operate with Brachyury in this function (Oliveri *et al.*, 2006). However, some genes that are known to be repressed by FoxA in this lineage did not show up in the differential transcriptomic data of this work, like *Gcm*, the main activator of mesodermal genes in SMC lineage after it receives input from Delta, and other genes downstream of Gcm (Oliveri *et al.*, 2006). This can be explained by the incomplete absence of *FoxA*, as detected by *in situ* hybridization. The *in situ* hybridization experiment of *brachyury* and *FoxA* in Brachyury perturbed embryos at 27hpf shows that there are still traces of *FoxA* transcripts in the vegetal plate (Figure 4.8C'). This presence of FoxA is probably still able to repress *Gcm*. Most probably, FoxA and Brachyury act together and drive the

repression of the mesodermal fate in the veg2 endoderm cells. There is some evidence that these two transcription factors might have a conserved co-operative role. Brachyury is expressed in the notochord (mesoderm) in the chordate clade, and it has been shown that it acts in tandem with FoxA in *Ciona*. (Katikala *et al.*, 2013; José-Edwards *et al.*, 2015). It has been shown that some Brachyury and FoxA targets contain binding motifs for both genes and that the occupancy of both domains is necessary to activate those targets (Katikala *et al.*, 2013; José-Edwards *et al.*, 2015).

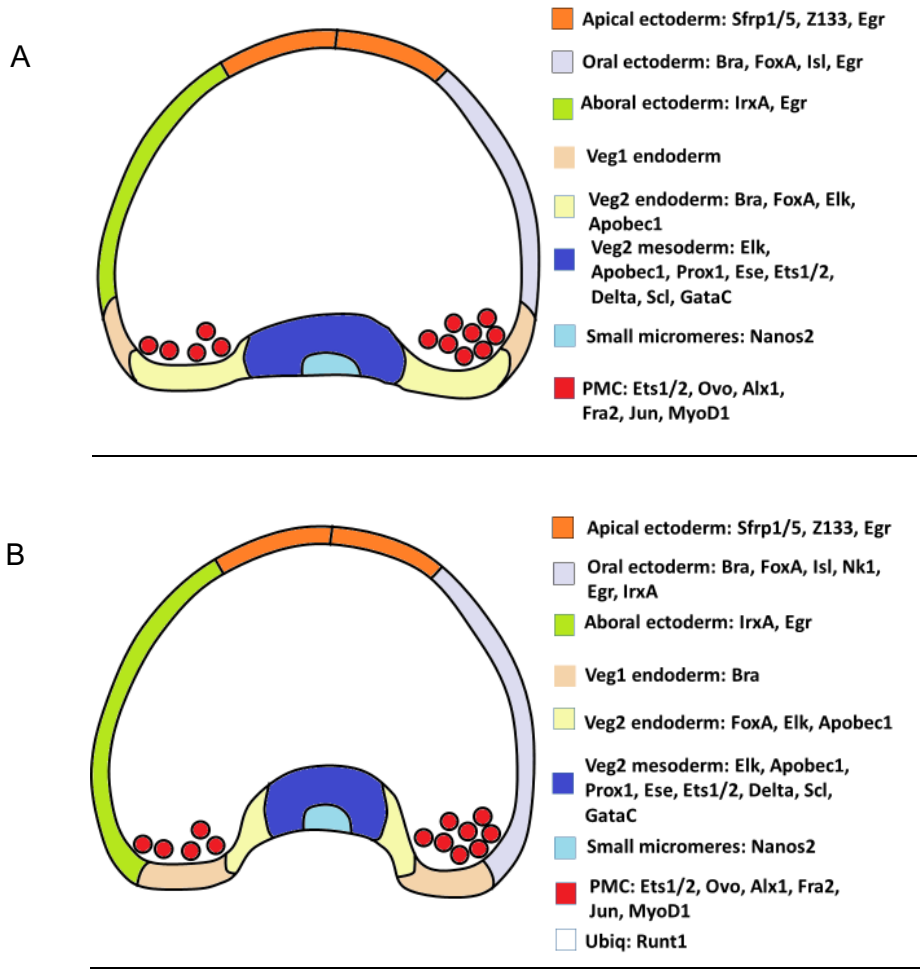


Figure 4.12 Graphical representation of 24h mesenchyme blastula and 27hpf early gastrula domains of expression of putative Brachyury targets in *S. purpuratus*. Ectodermal, mesodermal and endodermal specific genes are shown in their normal domains of expression.

Based on the described gene expression temporal and spatial patterns, GRNs models for both 24 and 27hpf endoderm is proposed in the last subsection of this chapter.

Brachyury targets at 48hpf

Genes under putative control of Brachyury in the ectoderm. There are 7 potential Brachyury targets in the oral ectoderm at 48hpf, on which Brachyury acts as an activator and 4 genes on which Brachyury acts as a repressor.

The down-regulated genes include the already described *FoxA* and *Isl*; then *Brn1/2/4*, *FoxABL*, *Pitx2*, *Ap2*, and *Klf13*.

Even though *FoxA* was not detected as differentially expressed at this stage, it is still considered a Brachyury direct target at 48hpf. Its expression probably recovers during gastrulation, even in the absence of Brachyury due to some other unknown mechanisms. However, based on the co-expression pattern with *Brachyury* at this stage (Figure 4.13) and the strong evidence from the earlier stages, it is placed in the oral ectoderm and endoderm GRN around Brachyury.

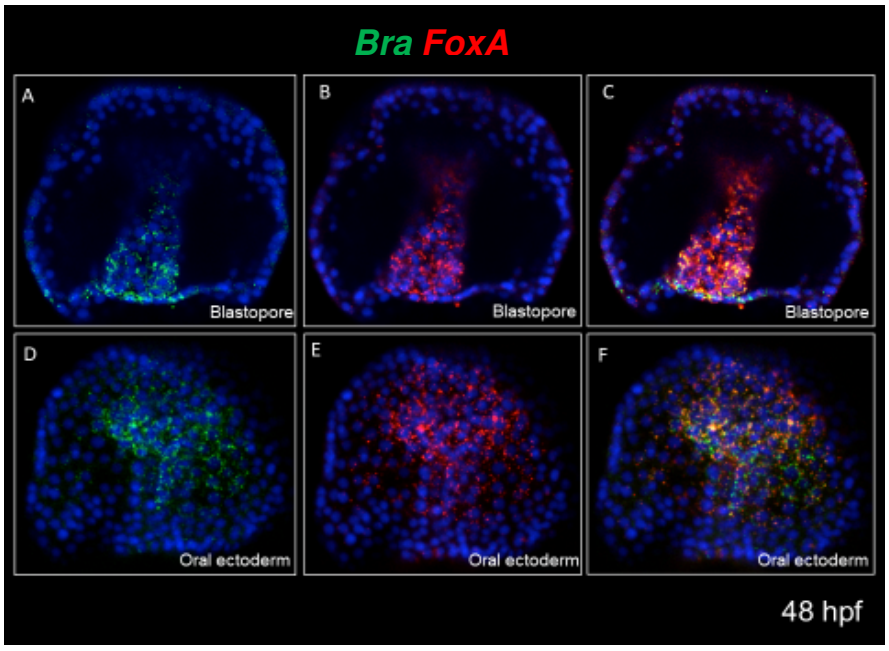


Figure 4.13 Double whole-mount fluorescent *in situ* hybridization of *Bra* and *FoxA* at 48hpf in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and co-expresses with *FoxA*, which is represented in red, in the ring of cells surrounding the blastopore (hindgut) and in the oral ectoderm. *FoxA* has another domain of expression which is throughout the gut. A) and D) *Brachyury*, B) and E) *FoxA*, C) and F) merged. Courtesy of Periklis Paganos.

Isl is tested for co-expression with *brachyury* also at 48hpf at Arnone's lab. The same pattern to 27hpf is observed in the late gastrula stage. *Brachyury* and *Isl* co-express in some cells of the oral ectoderm (Figure 4.14A, B, C) and some hindgut cells (Figure 4.14D). *Isl* contains one ChIP/ATAC-seq peak in the intergenic region downstream of its TSS. Based on the differential expression detected at 48hpf, ChIP/ATAC-seq peak, and the co-expression analysis, it is considered to be a direct target of Brachyury also at 48hpf, which seems to be an activator of *Isl*.

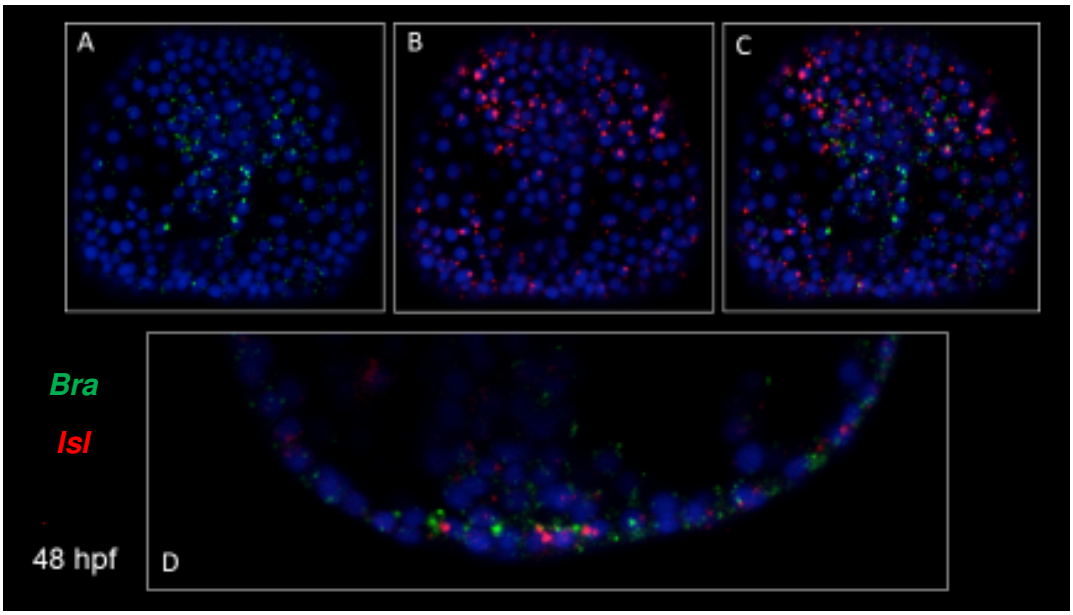


Figure 4.14 Double whole-mount fluorescent *in situ* hybridization of *Bra* and *Isl* at 48hpf in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and co-expresses with *Isl (islet),* which is represented in red, in the region of oral ectoderm (C), and the ring of cells surrounding the blastopore (D). A) *Bra*; B) *Isl*; C) and D) merged. Courtesy of Periklis Paganos.

Brn1/2/4 is a transcription factor, previously described as a midgut regulator in endodermal patterning (Yuh, Dorman and Davidson, 2005). However, later studies have shown that its expression is present in three domains, one endodermal in the foregut and two ectodermal – oral ectoderm and ciliary band (Cole and Arnone, 2009; Perillo *et al.*, 2018). *Brachyury* seems to co-express with *Brn1*/2/4 in the oral ectoderm region, as described in Figure 4.15. Since no ChIP/ATAC-seq Brachyury peaks were found in proximity to this gene's genomic location, it can be inferred that it is an indirect target.

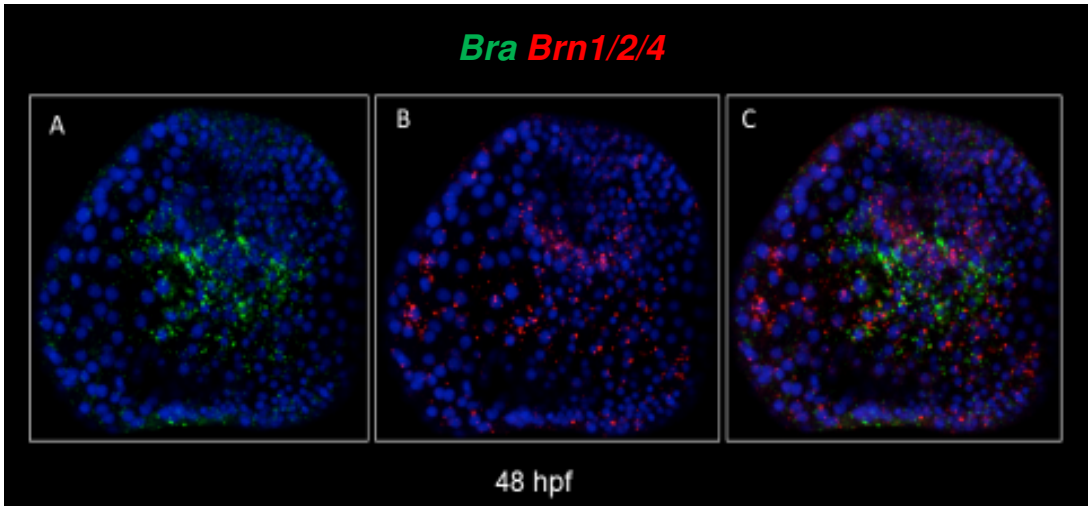


Figure 4.15 Double whole-mount fluorescent *in situ* hybridization of *Bra* and *Brn1*/2/4 at 48hpf in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and co-expresses with *Brn1/2/4,* which is represented in red, in the region of oral ectoderm. A) *Brachyury* expressed in the oral ectoderm; B) *Brn1/2/4* expressed in the oral ectoderm and ciliary band; C) *Bra* and *Brn1/2/4* are co-expressed in the oral ectoderm. Courtesy of Periklis Paganos.

FoxABL (SpFoxAB-like) is a Forkhead transcription factor, which belongs to the ancestral FoxAB family (not to be confused with the FoxA family). This family of Forkhead TFs has been lost in vertebrates and urochordates (Paps, Holland and Shimeld, 2012). *FoxABL* was first described in sea urchin as being zygotically expressed and without a specific localization during development by Tu *et al*, 2006. However, when assessed by *in situ* hybridization in this work, *FoxABL* transcripts were detected in the late gastrula stage as oral-ectoderm specific, where they co-express with *brachyury* (Figure 4.16C). At the 24hpf stage, no transcripts were detected via *in situ* hybridization. (Figure 4.16A).

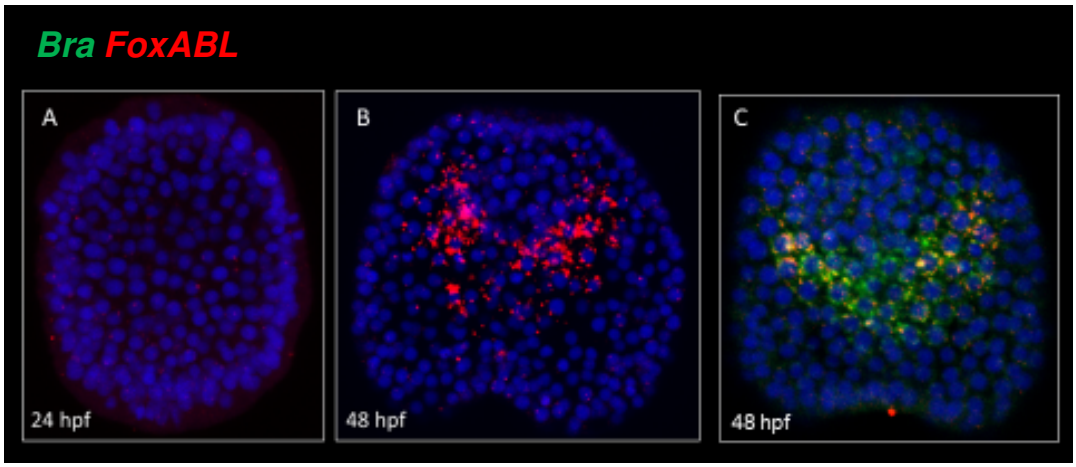


Figure 4.16 Single and Double whole-mount fluorescent *in situ* hybridization of *FoxABL* and *Bra* at 24hpf and 48hpf in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and *FoxABL* is represented in red. A) *FoxABL* transcripts are not detected in mesenchyme blastula stage. B) Expression of *FoxABL* is restricted to oral ectoderm C) *Brachyury* and *FoxABL* co-express in oral ectoderm domain. Courtesy of Periklis Paganos.

Another direct candidate that showed downregulation in the 48hpf transcriptomic dataset and has peaks around it in the 48hpf ChIP-seq datasets is a TF *Ap2* (Slota, Miranda and McClay, 2019). Homeobox-PRD TF *Pitx2*, normally expressed in the ectoderm near the tip of the archenteron (Duboc *et al.*, 2005; Luo and Su, 2012), seems to be an indirect target. Another candidate that Brachyury indirectly affects is zinc-finger TF *Klf13* expressed in the oral and apical ectoderm (Materna *et al.*, 2006).

There were 4 upregulated genes detected in the 48hpf differential dataset: the already described *Bra*, then *Dach*, *Not*, and *FoxQ2*.

Two upregulated genes are considered to be direct targets, based on the presence of ChIP-seq peaks, on which Brachyury acts as a repressor and blocks their expression in the oral ectoderm domain – TF *Dach* and TF *Not*. *Dach* (*dachshund*) is a member of the ski-sno family that is associated with Smad proteins in order to prevent the anti-proliferative effects of TGFβ signaling, expressed in the veg1 cells in blastula and the gastrula throughout the gut (Howard-Ashby, Stefan C. Materna,

et al., 2006). *Not* is normally expressed in the oral ectoderm after 30hpf (Li, Materna and Davidson, 2012). It might be the case that Brachyury is excluding it from the endoderm region in the late gastrula. These direct interactions are supported by the ChIP/ATAC-seq peaks present in their proximity.

Brachyury shows a high upregulation rate in the 48hpf RNA-seq dataset; still, no ChIP-peaks are detected close to its TSS. However, based on its strong upregulation and the peaks present in the 24hpf ChIP/ATAC-seq data sets, it is most likely that it auto-regulates its own expression in the late gastrula stage in both domains of expression.

Moreover, few other genes are considered to be Brachyury indirect targets based on the differential RNA-seq data. They include ciliary band specific genes: TF *Scratch*, TF *RasO*, and the receptor *Thytrprh* described in *L. variegatus* (Slota, Miranda and McClay, 2019); *Ngn* (Perillo *et al.*, 2018; Wood *et al.*, 2018), *Pitx2* described in *P. lividus* (Duboc *et al.*, 2005; Luo and Su, 2012), *Hnf6* (Otim *et al.*, 2004), *Isl* (this work and Perillo *et al.*, 2018), *Brn1/2/4* (this work and Cole and Arnone, 2009; Perillo *et al.*, 2018). Also, the aboral ectoderm genes that were affected are downregulated enzymes *Spec1* and *Spec2d*, mentioned before (Hardin *et al.*, 1985), and TF *Dlx* (Howard-Ashby, Materna, *et al.*, 2006); and upregulated TF *Klf2*/4 (Materna *et al.*, 2006). Apical ectoderm downregulated TFs were *FoxD* (Tu *et al.*, 2006) and *Klf13* (Materna *et al.*, 2006). It seems that Brachyury is able to exclude *FoxQ2* from the border between the apical and oral ectoderm since the apical ectoderm specific *FoxQ2* was upregulated in the RNA-seq dataset (Tu *et al.*, 2006). However, this interaction is probably mediated through another gene because no ChIP/ATAC-seq peaks were detected close to *FoxQ2*.

These results confirm Brachyury's importance in the ectoderm differentiation, as the results from 27hpf have already shown.

Genes under putative control of Brachyury in the mesoderm. At 48hpf, the number of mesodermal genes detected as upregulated is decreased compared to the 27hpf dataset. *Scl*, *Ets1/2*, *Ovo*, *Jun*, *MyoD1* still show upregulation (Table 4.2). Some additional mesodermal genes were affected as well: the PMC specific *Alx1*, *Nk7,* and *Tbr* (Oliveri, Carrick and Davidson, 2002; Ettensohn, 2003; Rafiq *et al.*, 2014; Dylus

et al., 2016), and the SMC specific *Erg* and *FoxY* (Rizzo *et al.*, 2006; Materna and Davidson, 2012; Andrikou *et al.*, 2013; Materna, Swartz and Smith, 2013). Despite the high presence of accumulated mesodermal transcripts, only *Scl* was found to contain one ChIP/ATAC-seq peak downstream of its TSS at 48hpf. It is still unclear how Brachyury, at that time point expressed only in oral ectoderm and midgut/hindgut domains of the embryo, would influence SMC specific *Scl* directly. The exclusion of *Scl* is happening under Brachyury's influence either in the oral ectoderm or, most likely, in the endoderm, where its regulatory state is suppressed since the blastula stage.

Most probably, other mesodermal genes were activated earlier, before the 48hpf, or in the proliferated mesodermal cells visible in the morphologically different Brachyury perturbed embryos. An alternative explanation of the upregulation of many PMC specific genes is SMC cells' ability to transfate into PMCs. It has been shown that in the case of embryos depleted of PMCs during development, SMC cells start to proliferate, and some of them change their fate to become PMCs to replace the depleted cells and to rescue the normal development (McClay and Ettensohn, 1988).

At 48hpf, 13 endodermal TFs are detected as downregulated. Most of them are expressed in the hindgut during normal development. They are *Cdx* (Arnone *et al.*, 2006; Annunziata and Arnone, 2014; Annunziata *et al.*, 2014), *Nkx6.1* (Arnone, unpublished and Israel *et al.*, 2016), *Isl* (this work and Annunziata *et al.*, 2014; Perillo *et al.*, 2018), *FoxD* (Tu *et al.*, 2006), *Osr* (Materna *et al.*, 2006), which also contain ChIP-seq peaks around them at 48hpf; then *Isx* (Arnone, unpublished and Annunziata *et al.*, 2014), *Hb9,* described in *P. lividus* (Bernardo *et al.*, 2000), *FoxI* (Tu *et al.*, 2006), signaling molecule *Wnt3* (McClay, Miranda and Feinberg, 2018), and *Asb5* (Zazueta-Novoa and Wessel, 2014) that are without ChIP-seq peaks. A midgut terminal differentiation gene *ManRC1* is observed too. The eight already described TFs were found to be also expressed in the foregut, and they were downregulated: *Isl*, *Wnt3*, *FoxD*, *SoxE*, Pitx2, and *Unc4*.1 (Howard-Ashby, Stefan C Materna, *et al.*, 2006). *Krl* is a transcription factor that is expressed in the endoderm before the gastrulation starts, and after gastrulation, its expression diminishes (Stepicheva et al., 2015; Materna et al., 2006; Minokawa et al., 2004; Howard et al.,

2001). It could be inferred that Brachyury can repress its endodermal fate in the endoderm at the late gastrula, and therefore it is considered to be an endodermal target in the hindgut and midgut.

Previously described stomach-specific marker already present at the late gastrula stage in the midgut (Figure 18A), *Macrophage mannose receptor*, *ManRC1A* (Annunziata, 2011; Annunziata *et al.*, 2014) was detected as downregulated in Brachyury perturbed dataset at 48hpf. Since the gut in perturbed embryos showed malformations compared to the wild-type, this gene was tested by *in situ* hybridization. This is a gut terminal differentiation gene that showed complete absence after the perturbation of Brachyury (Figure 4.17A'). However, *ManRC1A* does not appear in the ChIP-seq data set, and there is no Brachyury binding site in its proximity; therefore, it can be an indirect Brachyury target.

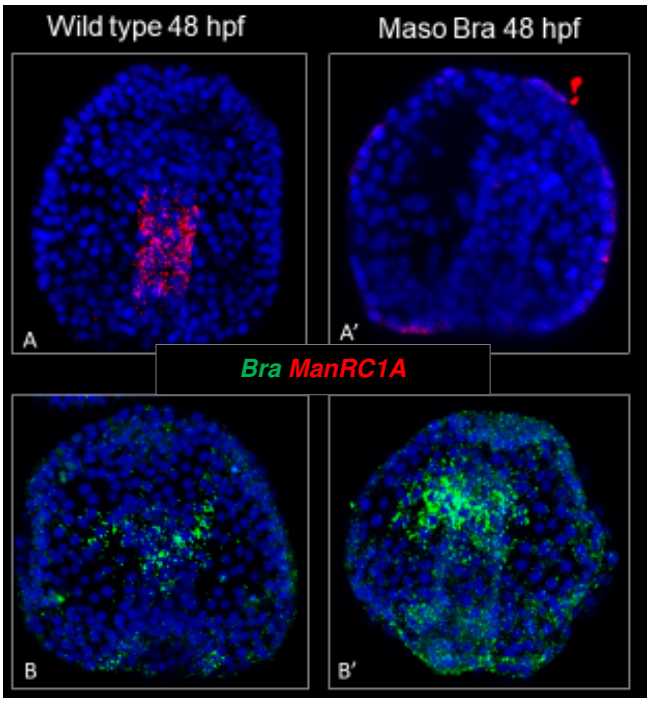


Figure 4.17 Whole-mount fluorescent *in situ* hybridization of *mannose receptor* and *brachyury* at 48hpf in normal and Brachyury perturbed condition in *S. purpuratus*. *Brachyury* (*Bra*) is represented in green, and *mannose receptor* (*ManRC1A*) is represented in red. A) and B) wild type condition; A') and B') Brachyury morpholino injected embryos. A) *Mannose receptor* is normally expressed in the midgut. A') after Brachyury knock-down the expression of *Mannose receptor* is completely abolished. B) oral ectoderm and hindgut expression of *Brachyury* transcripts in wild type condition. B') Increase in the number of *Brachyury* transcripts in the oral ectoderm and hindgut. Courtesy of Periklis Paganos.

As already mentioned, oral ectoderm specific TF *Not* shows upregulation in the 48hpf Brachyury knock-down dataset, and it contained ChIP-peaks in its proximity. It has been noticed that, even though *Not* is detected in blastula in the endodermal lineages, its expression diminishes in this domain in gastrula, where it stays active only in the oral ectoderm region (Li, Materna and Davidson, 2012). Therefore, it could be a direct target that Brachyury probably represses in the hindgut.

One peculiar situation is observed with the *SoxE* transcription factor. *SoxE* is expressed at the tip of the archenteron, and it is of mesodermal origin. Nevertheless, it does have a Brachyury ChIP-seq peak in its exon. There has been some evidence that Brachyury might have its expression activated in the late gastrula stage in the foregut, particularly in the SMC derived cells. When the *brachyury* gene was first isolated in *S. purpuratus*, the initial result showed it to be present in the secondary mesenchyme founder cells (Harada, Yasuo and Satoh, 1995; Peterson, Harada, *et al.*, 1999). Although Rast *et al*. later disproved this (Rast *et al.*, 2002), it could be possible that at the precise moment, just before the stomodeum is formed, an unknown signal from the oral ectoderm turns on the expression of *brachyury* in those cells. It could be that this amount of mRNA is low, and it requires a specific timing to be captured by *in situ* hybridization. This idea is also supported by the observed localization of SpBra protein in the late gastrula *S. purpuratus* embryo with immunohistochemistry (Figure 4.18).

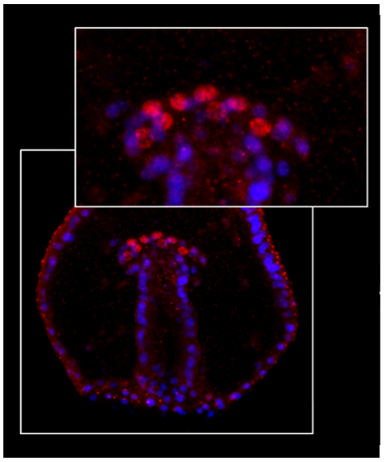


Figure 4.18 Localization of the Brachyury protein in the developing coelomic pouches of the late gastrula *S. purpuratus* embryo. Nuclei are stained with blue DAPI, Brachyury positive nuclei are red.

Ubiquitous TFs and signaling molecules affected by Bra knock-down based on the 48hpf differential expression data. Ubiquitously expressed *Runt1* (Coffman *et al*., 1996, 2002), with its cofactor *Cbfb* (Robertson *et al*., 2008) and *Erf* (Rizzo *et al*., 2006) were downregulated.

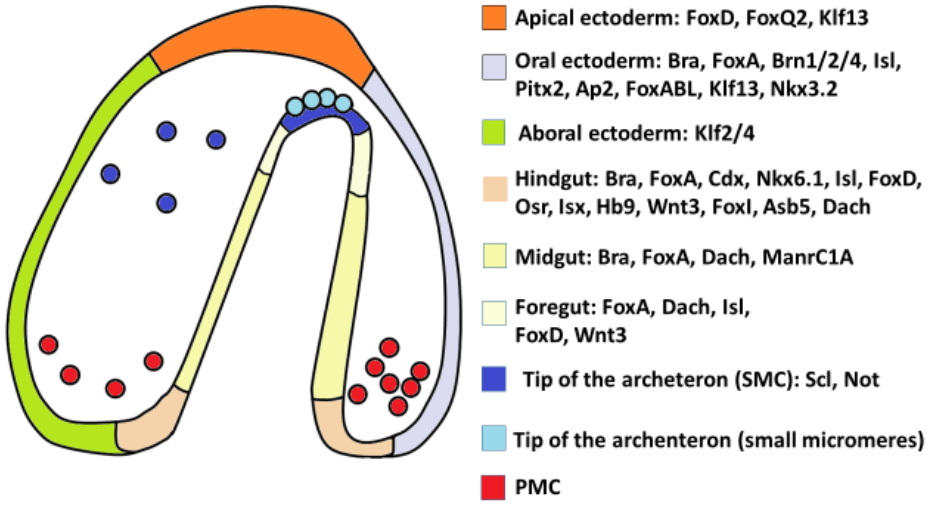


Figure 4.19 Graphical representation of 48hpf late gastrula domains of expression of putative Brachyury targets in *S. purpuratus*. Ectodermal, mesodermal and endodermal specific genes are shown in their normal domains of expression.

Based on the described temporal and spatial gene expression patterns, differential RNA-seq analysis, as well as combined ChIP-seq and ATAC-seq data, GRNs models for oral ectoderm and gut expression domains are proposed in the subsequent section.

Even though a large number of TFs and signaling molecules were affected by Brachyury perturbation, the proposed GRNs are mostly composed of the targets which were found to share the same spatial domains with Brachyury. However, some genes are placed as inactive in the two domains of expression, oral ectoderm or the endoderm, based on the ChIP-seq datasets and strong upregulation after the injection of Brachyury MASO.

4.2.7 Reconstruction of the GRN around Brachyury

One of the largest and the frequently updated Gene Regulatory Network models are sea urchin GRNs available at http://www.echinobase.org/endomes/. They represent the cumulative knowledge gathered from different research groups involved in the evo-devo echinoderm studies. It serves as a reference point - it is validated and can serve as an interaction prediction model for developmental biology (Davidson, 2010). The GRNs are currently available for ectoderm and endomesoderm domain of expression from 0-30hpf. The GRNs are represented as hierarchical structures that are interconnected via signaling events.

As described previously, the expression of *Bra* starts first in the endomesoderm lineage, around 15hpf. At this time point, endomesoderm still has the regulatory state of a single tissue. At about 21hpf, this regulatory state is differentiated between the two tissues that the endomesoderm linage produced – the veg2 mesoderm and the veg2 endoderm. At the same time, *Bra* starts to be transcribed in the oral ectoderm as well. In veg2 mesoderm, the endoderm fate is repressed, while in veg2 endoderm, the mesoderm fate is repressed. The veg2 mesodermal fate is governed by the Delta/Notch signaling and *Gcm*, while it has been shown that *FoxA* is the main actor in repressing the mesodermal fate in veg2 endoderm as discussed previously in this chapter (Oliveri *et al.*, 2006). After the invagination of the veg2 endoderm cells to form the gut, *Bra* stops to be transcribed, and its transcription starts in the veg1 endoderm lineage (around 24-25hpf). Therefore, to correctly reconstruct the GRN, Brachyury's dynamic expression has to be taken into account. At 48hpf endodermal protein, Brachyury localization is confined to the hindgut, midgut, and oral ectoderm.

Gene regulatory networks were drafted using the *BioTapestry* software, where each node represents a gene or signaling molecule involved in an interaction. Interactions can be positive (activating) or negative (repressing). During the late blastula and early gastrula stage, 18 genes were selected to add to the existing and published GRN around Brachyury in the endoderm. Oral ectoderm GRN was reconstructed at 27hpf by adding 4 nodes. The known upstream regulators of Brachyury and the known downstream targets in the endomesoderm and ectoderm GRNs are published on http://www.echinobase.org/endomes/, and they were used as a GRN

template. At 48hpf stage endoderm GRN, published by Annunziata *et al.*, was reconstructed by adding 12 additional nodes (Annunziata and Arnone, 2014; Annunziata *et al.*, 2014). Oral ectoderm at 48hpf is constructed *de novo*.

Reconstruction of the GRN around Brachyury in the oral ectoderm

The expression of *brachyury* is controlled by Nodal signaling. It has been shown that Nodal signaling is necessary for the specification of the oral ectoderm on the ventral side of the embryo, and it affects *brachyury* expression directly by activating it (Duboc *et al.*, 2004; Lapraz, Besnardeau and Lepage, 2009). FoxA input on *Bra* was inferred based on the *FoxA* MASO injection that resulted in *Bra's* absence in the oral ectoderm in *P. lividus* (Saudemont *et al.*, 2010). Published oral ectoderm (stomodeum specific) GRN up to 30hpf seemed to be very shallow; however, based on the data collected in this work, it can be inferred that Brachyury is one of the main players, controlling in total 14 genes from 27-48hpf. At the early gastrula stage, it seems that Brachyury has a positive input on 3 TFs – *FoxA*, which has been shown before, *Isl*, and *Nk1*. This is based on the presence of ChIP-seq Brachyury peaks around those genes. Moreover, Brachyury can regulate its own expression directly. *Runt1*, a ubiquitously expressed gene during the sea urchin development, was upregulated after the injection of Bra MASO, and it contains the Bra ChIP-seq peak, which shows that it is probably under the direct influence of Brachyury. Most likely, Bra is regulating the levels of expression of *Runt1* by repressing it. *Elk*, also ubiquitously expressed during development, seems to be excluded from the ectoderm in the late blastula (Rizzo *et al.*, 2006), most probably through an intermediate signaling event, since no Bra ChIP-seq peaks were found in its proximity. The 27hpf oral ectoderm GRN is represented in Figure 4.20.

At 48hpf, the involvement of Brachyury in oral ectoderm specification seems to be even larger. It activates directly Brn1/2/4, *Isl*, *FoxA* and *Ap2*, and indirectly *Pitx2*, *FoxABL* and *Klf13*. Moreover, it seems that it represses directly the expression of *Not*, which is normally expressed in the oral ectoderm, and *Dach*, which is expressed in the archenteron. It seems that Brachyury also defines the boundary between the oral ectoderm and apical ectoderm by indirectly repressing *FoxQ2*.

Reconstruction of the GRN around Brachyury in the endoderm

GRNs in the endodermal domains are reconstructed in 3 developmental time points – 24hpf, 27hpf, and 48hpf.

As mentioned before, Brachyury most likely directly represses the mesodermal fate in the veg2 endodermal cells. This is supported by the presence of ChIP-seq Brachyury peaks near the differentially expressed genes after Bra MASO: *Delta*, *Spalt*, *Scl*, *Ese*, and *GataC*. It indirectly represses another SMC specific TF, *Prox1*. It seems that it also has the ability to repress the endodermal expression of *Egr*, which is expressed only in the ectoderm, and it regulates the levels of *Runt1*. Moreover, 6 PMC specific genes appeared to be upregulated by the Bra MASO; direct targets are probably *Ovo*, *Fra2*, and *Jun*, indirect *Alx1*, *Ets1/2*, and *MyoD1*. Brachyury is directly activating *FoxA* and *Otx*, and indirectly *Apobec1*. Brachyury itself is activated by the inputs from Beta-catenin, *GataE*, and *Hox11*/13b. See Figure 4.22.

As the veg2 cells invaginate and form the blastopore at 27hpf, Brachyury is activated in the veg1 endodermal cells. In this domain, Brachyury is activated by the Beta-catenin and *Hox11/13b*, and then it directly activates *Isl*, *Nk1*, *Hnf1*, and *Otx*.

Since Brachyury shows a dynamic expression pattern, it continues to be expressed around the blastopore, and at the time of reaching the late gastrula stage, this domain of expression is considered to be the presumptive hindgut. It seems that in the hindgut, Brachyury is directly activating hindgut specific genes – *Cdx*, *Nkx6.1*, *Osr*, and *Isl* (also expressed in the presumptive foregut and oral ectoderm) and *FoxD*, also expressed in the presumptive foregut. Indirect hindgut targets are *FoxI*, *Asb5*, *Wnt3*, *Hb9*, and *Isx*. As a repressor, it seems that it can directly exclude *Not* expression, and indirectly, *Nkx3.2* (foregut specific) and *Krl*, expressed in the endoderm before gastrulation.

Based on the observation of the Brachyury protein presence in the presumptive midgut (veg1 endoderm descendent cells), Brachyury is placed as a node in this domain. Here, it is possible that it directly activates *FoxA* and indirectly *Mannose receptor* (*ManRC1A*), which is a terminal differentiation gene. Moreover, in this

domain, it is also possible that it excludes the expression of *Not*, *Krl*, and *Nkx3.2*. For the 48hpf endoderm GRN, refer to Figure 4.23.

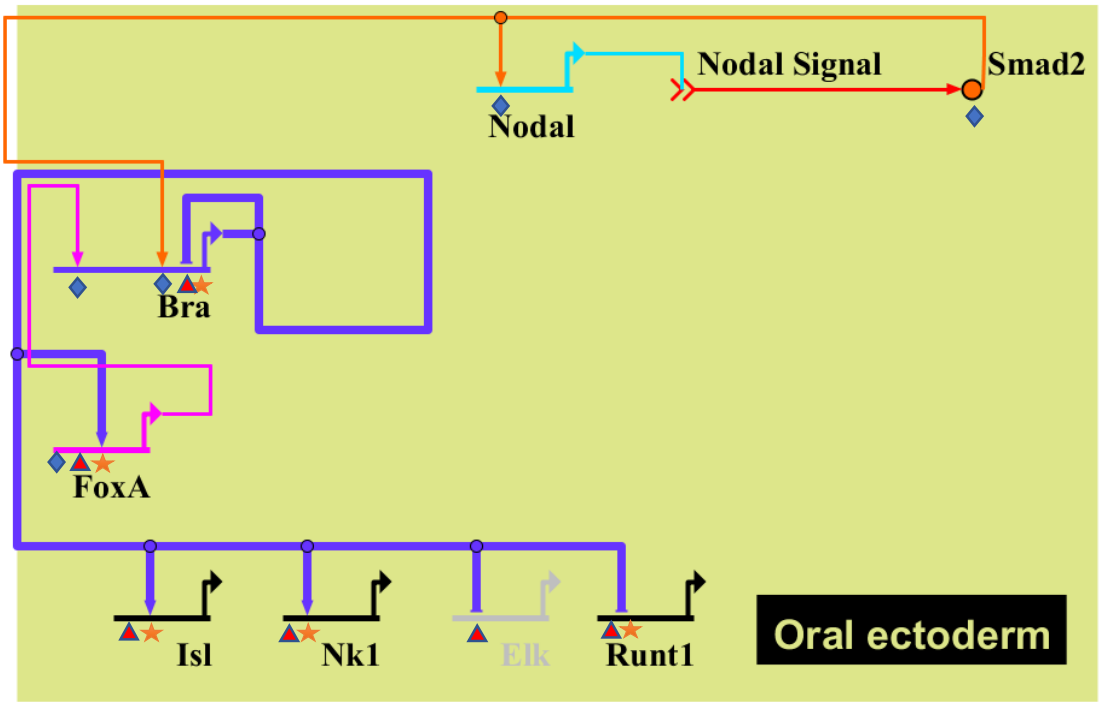


Figure 4.20. Schematic representation of the Oral ectoderm Gene regulatory network around Brachyury at 27hpf in *S. purpuratus* using BioTapestry.

(http://www.biotapestry.org/) Inputs coming from the same gene are shown. The arrows represent positive inputs. The horizontal bar represents negative inputs. The gray colored genes are inactive. Blue diamonds indicate the interactions based on published oral ectoderm 21-30hpf GRN (http://www.echinobase.org/endomes/#EctodermNetwork). Red triangles indicate the expression data quantified by RNA-seq after Bra MASO. Orange stars indicate direct inputs based on Brachyury ChIP-seq data.

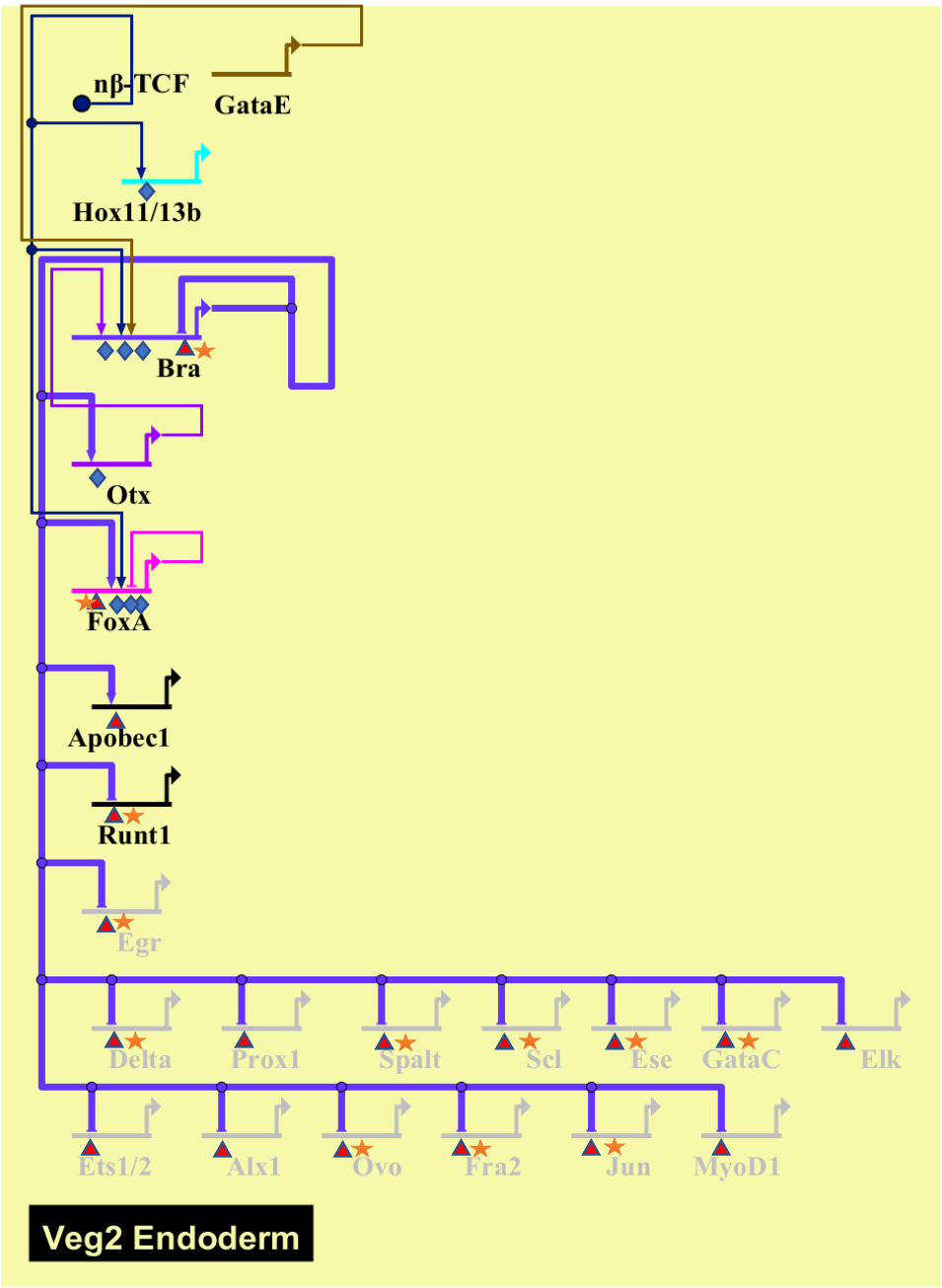


Figure 4.21 Schematic representation of the Veg2 endoderm Gene regulatory network around Brachyury from 21 to 24hpf in *S. purpuratus* using BioTapestry.

(http://www.biotapestry.org/) Inputs coming from the same gene are shown with the unique colored lines. The arrows represent positive inputs. The horizontal bar represents negative inputs. The gray colored genes are inactive. Blue diamonds indicate the interactions based on published endomesoderm 21-30hpf GRN (http://www.echinobase.org/endomes/#Veg-21-30-NetworkDiagram). Red triangles indicate the expression data quantified by RNA-seq after Bra MASO; Orange stars indicate direct inputs based on Brachyury ChIP-seq data.

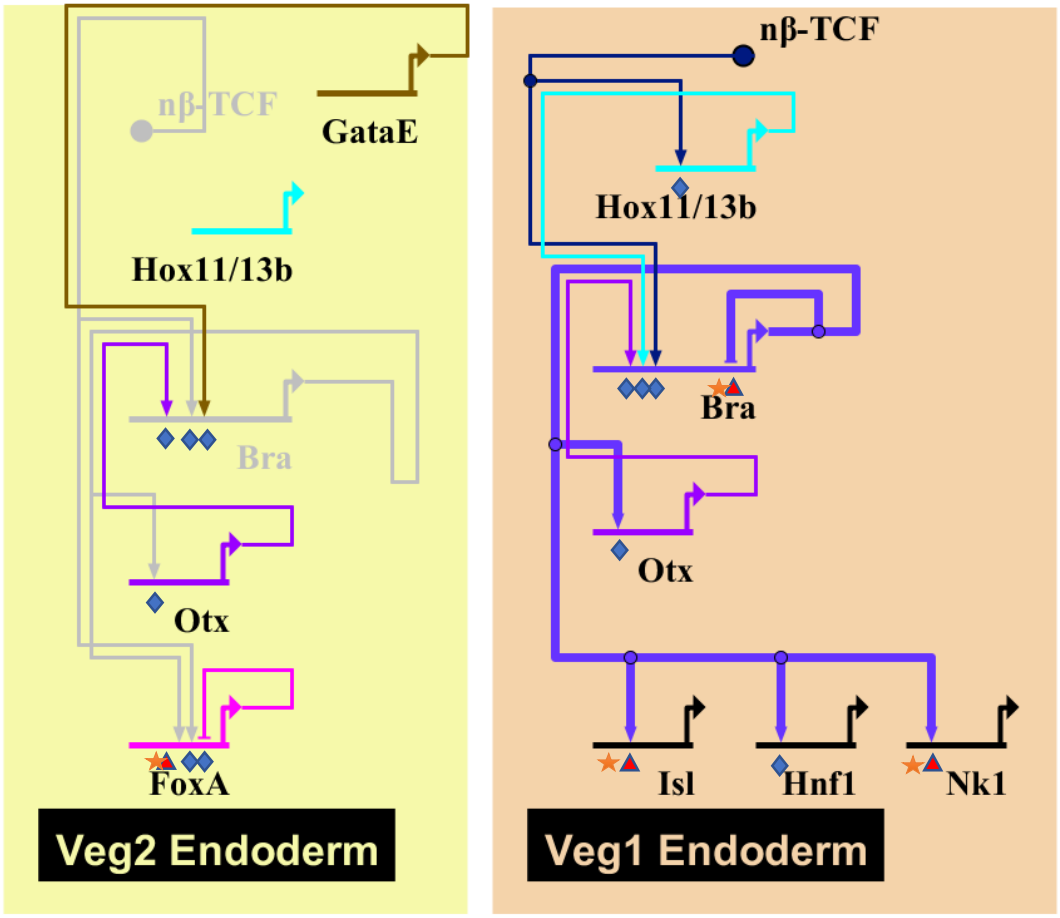


Figure 4.22 Schematic representation of the Veg2 and Veg1 endoderm Gene regulatory networks around Brachyury at 27hpf in *S. purpuratus* using BioTapestry.

(http://www.biotapestry.org/) Inputs coming from the same gene are shown with the unique colored lines. The arrows represent positive inputs. The horizontal bars represent negative inputs. The gray colored genes are inactive. Blue diamonds indicate the interactions based on published endomesoderm 21-30hpf GRN (http://www.echinobase.org/endomes/#Veg-21-30-NetworkDiagram). Red triangles indicate the expression data quantified by RNA-seq after Bra MASO. Orange stars indicate direct inputs based on Brachyury ChIP-seq data.

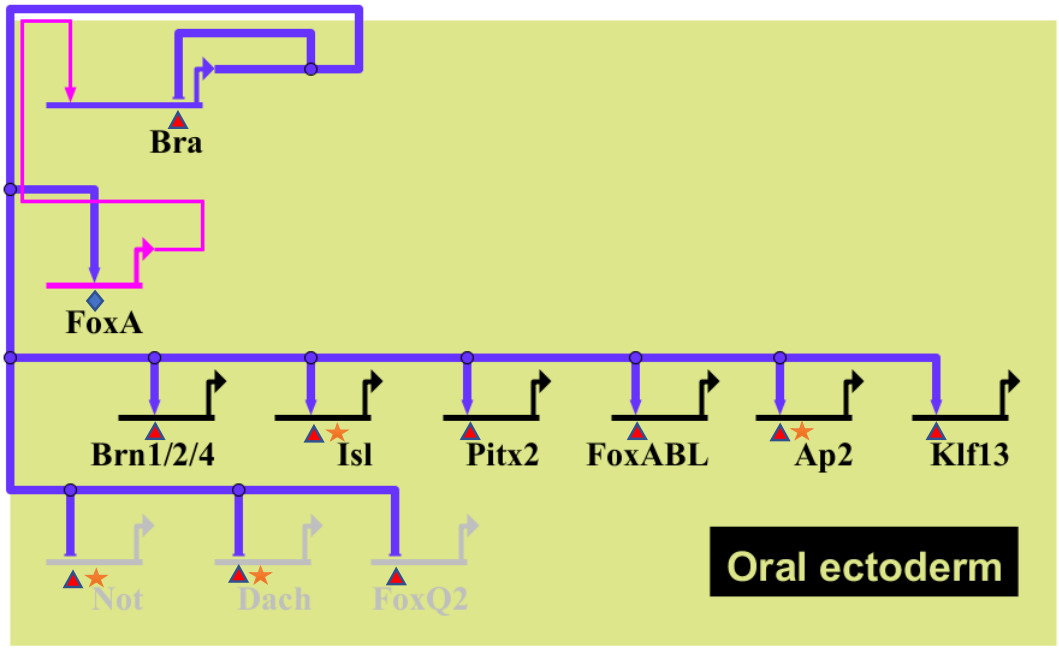


Figure 4.23 Schematic representation of the Oral ectoderm Gene regulatory network around Brachyury at 48hpf in *S. purpuratus* using BioTapestry.

(http://www.biotapestry.org/) Inputs coming from the same gene are shown with the unique colored lines. The arrows represent positive inputs. The horizontal bars represent negative inputs. The gray colored genes are inactive. Blue diamonds indicate the interactions based on published oral ectoderm 21-30hpf GRN (http://www.echinobase.org/endomes/#EctodermNetwork). Red triangles indicate the expression data quantified by RNA-seq after Bra MASO. Orange stars indicate direct inputs based on Brachyury ChIP-seq data.

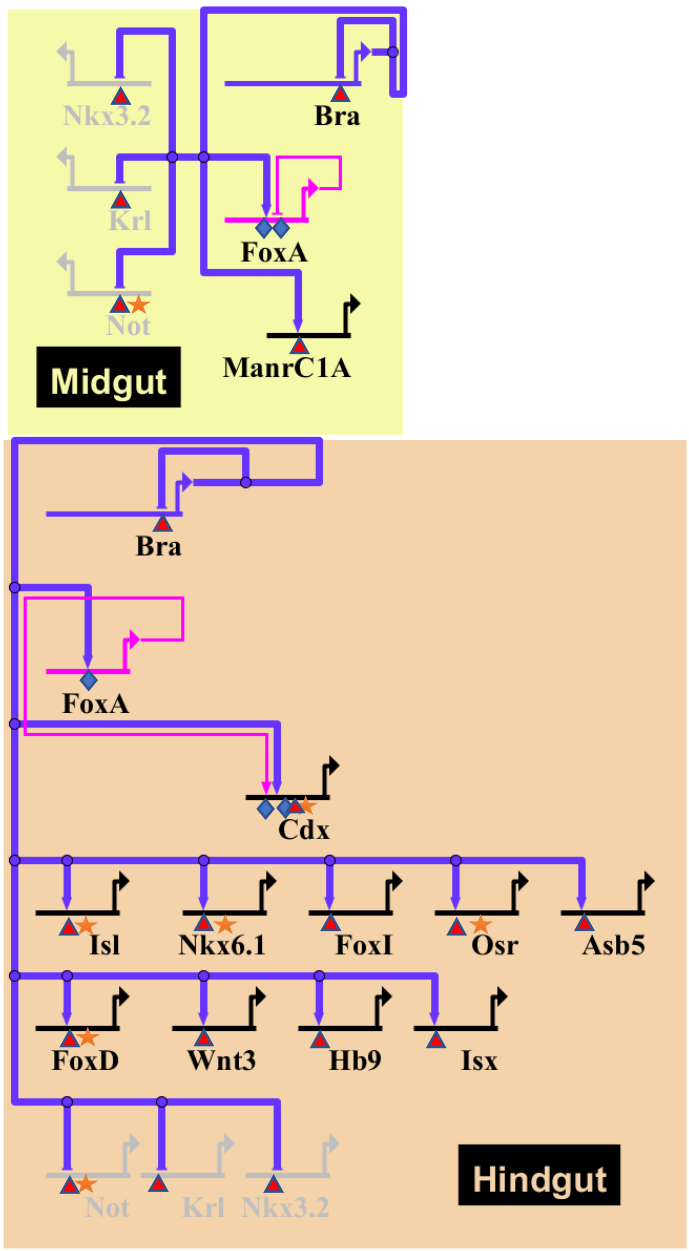


Figure 4.24 Schematic representation of the endoderm Gene regulatory network around Brachyury at 48hpf in *S. purpuratus* using BioTapestry.

(http://www.biotapestry.org/) Inputs coming from the same gene are shown with the unique colored lines. The arrows represent positive inputs. The horizontal bars represent negative inputs. The gray colored genes are inactive. Blue diamonds indicate the interactions based on published gut 48hpf GRN by Annunziata *et al.*, 2014. Red triangles indicate the expression data quantified by RNA-seq after Bra MASO. Orange stars indicate direct inputs based on Brachyury ChIP-seq data.

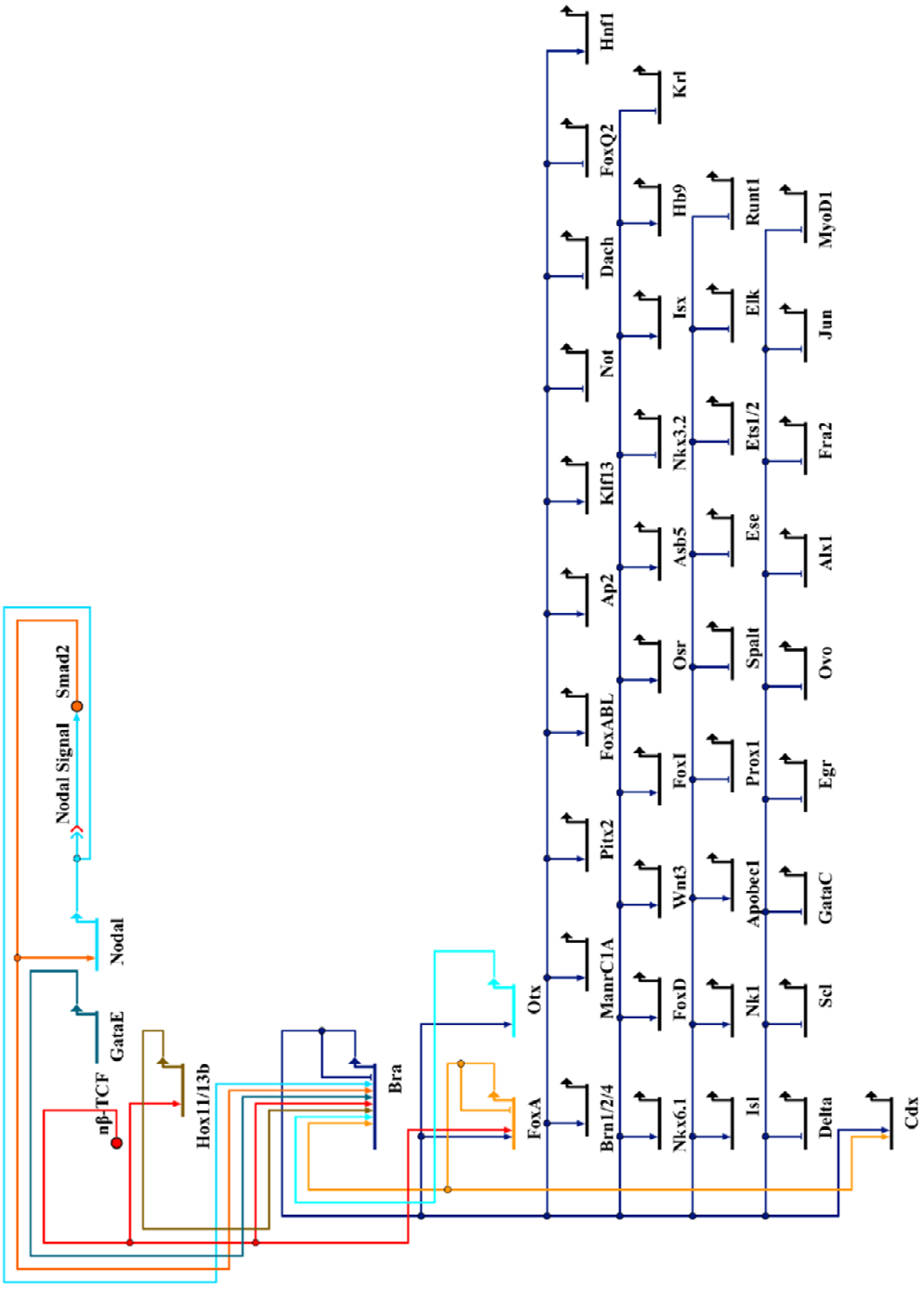


Figure 4.25 Schematic representation of the Gene regulatory network around Brachyury 24-48hpf in *S. purpuratus* using BioTapestry. View from the Genome.

(http://www.biotapestry.org/) All interactions upstream and downstream of Brachyury are represented. Inputs coming from the same gene are shown with the unique colored lines. The arrows represent positive inputs. The horizontal bars represent negative inputs.

These results show that the GRN around Brachyury is deep. Brachyury controls directly and indirectly 40 TFs/signaling molecules (Summarized in Figure 4.25). Each of the 20 direct and indirect targets activated by Bra, except the known terminal differentiation gene *ManrC1A* (Annunziata and Arnone, 2014), has their own sub-circuits that can control a large number of gene batteries. This study shows that Brachyury is one of the main regulatory genes able to establish ectodermal and endodermal regulatory states and, moreover, to exclude the alternative, mesodermal state in the endoderm lineage.

4.3 Conclusions

In this chapter, the main processes affected after Brachyury perturbation just before and after gastrulation are described. Gene ontology analysis showed that the most affected processes were related to metabolism and that the most affected proteins were the ones that have transporter activity, nucleic acid binding, and transcription factor activity. This work shows that the depletion of Brachyury TF leads to abnormal development that has many processes affected.

Brachyury ChIP-seq and ATAC-seq data were coupled with the transcriptomic analysis to reveal the consequences of the Brachyury absence and the regulatory states of a normal, healthy developing *S. purpuratus* embryo. To this aim, six main properties of Brachyury were discovered:

1. Before gastrulation, the important function of the Brachyury seems to the repression of mesodermal fate in the veg2 endodermal lineage, probably in tandem with the TF FoxA;

2. During the mesenchyme blastula stage, Brachyury acts as an activator in both oral ectoderm and vegetal plate

3. At the end of gastrulation, the most important function of Brachyury is to act as an activator in the hindgut, most probably with other endodermal transcription factors in order to “build” a functional tripartite gut;

4. At the gastrula stage, Brachyury is an important activator in the oral ectoderm, where it activates *FoxA* and leads to the opening of the mouth.

5. Brachyury has the ability to auto-regulate itself as a repressor in both domains of its expression

6. Moreover, it seems that Brachyury serves an important role in setting boundaries between apical ectoderm, aboral, and oral ectoderm tissues, and therefore it is probably involved in the dorsoventral patterning of the embryo.

The Gene Regulatory Network around Brachyury was reconstructed. Although many genes were proven to co-express with Brachyury, more detailed analysis will be required in the future, such as testing the putative CRMs identified through ChIP-seq/ATAC-seq combination.

Contribution statement

Dr. Maria I. Arnone performed MASO microinjections. Perklis Paganos, a PhD student from Dr. Arnone's lab, took care of the embryo cultures, imaged the embryos, and collected RNA for sequencing. P. Paganos performed whole-mount *in situ* hybridization experiments. The author of this thesis performed the literature search and RNA-seq computational analysis. ChIP-seq experiments were performed by Dr. Carmen Andrikou, a former PhD student from Dr. Arnone's lab. Dr. Danila Voronov, a Post-Doc from Arnone's lab, performed ChIP-seq computational analysis up to the peak calling. Marta Magri, a collaborator from Prof. JL Gómez-Skarmeta's lab, Dr. D. Voronov and Dr. Claudia Cuomo, Post-Docs from Arnone's lab and the author of this thesis, equally contributed in conducting the ATAC-seq experiments. M. Magri and Dr. D. Voronov performed computational analysis of the ATAC-seq sequenced reads up to the peak calling. The author of this thesis performed the ChIP-seq *de novo* motif analysis, filtering the ATAC-seq/ChIP-seq NarrowPeak datasets, and drafted the GRNs.

CHAPTER 5

DIFFERENTIAL GENE EXPRESSION ANALYSIS AFTER BRACHYURY KNOCK-DOWN IN *P. LIVIDUS*

This chapter shows the preliminary results of differentially expressed genes after the knock-down of Brachyury in the Mediterranean sea urchin species *Paracentrotus lividus.* Due to issues described in text below, no immediate safe conclusions could be drawn regarding Brachyury's role in *P. lividus.*

5.1 Introduction

Paracentrotus lividus is a sea urchin species that has been used for decades as a model organism in evolutionary and developmental biology. Its development is similar to that of other echinoid species, but compared to *S. purpuratus*, normal development occurs at a higher temperature (18 vs. 15 °C) and is much faster. For instance, the mesenchyme blastula stage is reached by 15hpf, the blastopore is open by 18hpf (whereas in *S. purpuratus*, it is open at 27hpf), and the late gastrula stage is reached at 24hpf, while in *S. purpuratus*, the same developmental stage is reached at 48hpf. To detangle the evolution of the transcription factor Brachyury, *Paracentrotus lividus* was chosen as a comparison species due to Brachyury protein's different expression pattern in the gastrula gut compared to *S. purpuratus* (Chapter 3) and its recently annotated genome. As previously described, *P. lividus* shows the same Brachyury mRNA pattern, as in the other sea urchin species tested, being expressed in the oral ectoderm and endoderm, in particular the ring surrounding the blastopore (Chapter 3). The newly assembled genome (unpublished, provided by the authors to Dr. M.I. Arnone lab that is part of the sequencing consortium for the *P. lividus* genome) gave the opportunity to look at the gene expression of Brachyury at a system-wide level. The use of *P. lividus* genome allowed to generate differential gene expression datasets after Brachyury

knock-down that can be used in the future to perform the omics analysis at the same level as described for S*. purpuratus* (Chapter 4).

5.2 Results and discussion

5.2.1 Brachyury knock-down and phenotype analysis of P. lividus *embryos*

To unravel Brachyury downstream targets in *P. lividus*, its translation was blocked via injecting zygotes with specific antisense morpholino oligonucleotides designed against the *PlBrachyury* mRNA. The wild-type and injected embryos at early gastrula (18hpf) and late gastrula (24hpf) stage were collected and sequenced. These developmental time points were chosen to be in accordance with the developmental time points selected for *S. purpuratus*.

Wild type embryo at 18hpf has a spherical shape, blastocoel filled with primary mesenchyme cells, and thickened vegetal plate that started to invaginate and form the blastopore. Injected embryos at 18hpf show delayed development, as they are still in the mesenchyme blastula stage without the formed blastopore. 24hpf wild type embryos were at the late gastrula stage, with the gut well elongated (Figure 5.1 white arrow) and the PMCs are re-arranged in the vegetal part of the embryo already forming skeleton spicules (Figure 5.1, green arrow). Injected 24hpf embryos showed high morphological variability, ranging from embryos that only started to gastrulate (Figure 5.1) to the ones that had archenteron fully developed (Figure 4.3B').

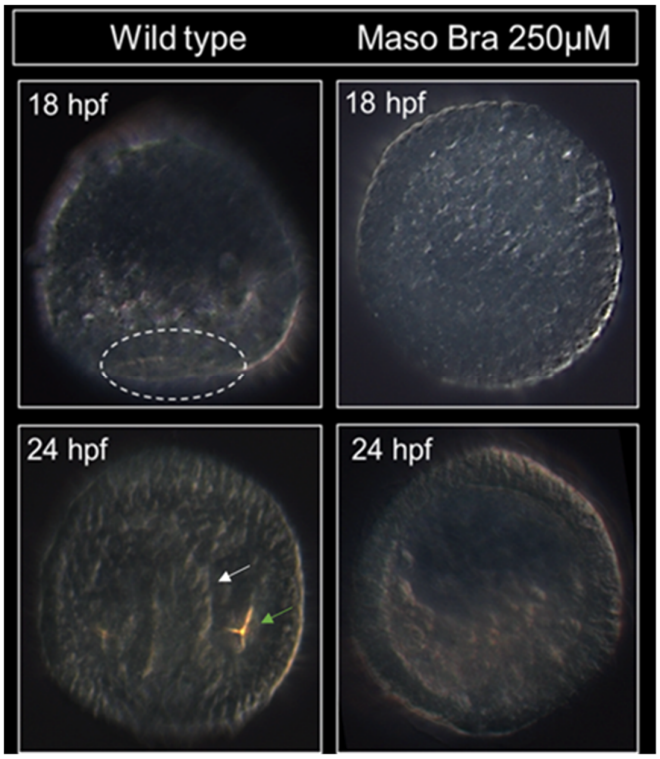


Figure 5.1 Phenotypic differences in control and Brachyury MASO treated embryos at early gastrula (18hpf) and late gastrula (24hpf) of *P. lividus*. The white dotted oval shape marks the blastopore opening. The white arrow in the 24hpf control embryos marks the archenteron. Green arrow marks forming skeleton spicules. 24hpf embryo is represented in the lateral view. Courtesy of Periklis Paganos.

Before testing for differential gene expression between the control and Brachyury knock-down embryos, all the samples were checked for the possible sources of variation using the Principal Component Analysis (PCA). The treatment was conducted in 4 biological replicates for each stage. After running the PCA and preliminary differential gene expression analysis using all four biological replicates for each stage, the number of differentially expressed genes was very low. The preliminary PCA plots showed that some treatment and control samples were very similar to each other (Figure not shown). It seems that some of the treated samples were more resilient to morpholino injection, and therefore they were not used in the final analysis. To select the samples at the early gastrula stage (18hpf) on which MASO had an effect, the transcript count tables were checked for the levels of expression of *FoxA*, a conserved Brachyury target (Rast *et al.*, 2002; Saudemont *et*

al., 2010; Peter and Davidson, 2011a; Katikala *et al.*, 2013; Ikeda and Satou, 2017). The Brachyury knock-down effect on *FoxA* expression in *P. lividus* was assessed by the *in situ* hybridization (Figure 5.4), which showed that *FoxA* is indeed downregulated, and therefore confirming the conserved regulatory relationship between *Bra* and *FoxA*. Based on this, treatment samples with a lower number of *FoxA* transcripts detected, compared to the wild type were selected for the analysis (Figure 5.2). However, at 18hpf, it seems that *FoxA* is not transcribed yet in the oral ectoderm region. It could be possible that the developmental timing was not in concordance with one of the early gastrulae of sea urchin (27hpf) or that in *P. lividus*, *FoxA* starts to be expressed later during development in this domain. For 24hpf, biological replicates that showed high similarity between the treatment and control were not considered for the differential analysis, and the differential gene expression analysis was performed using three biological replicates, removing the batch effects detected by the PCA analysis (Figure 5.3).

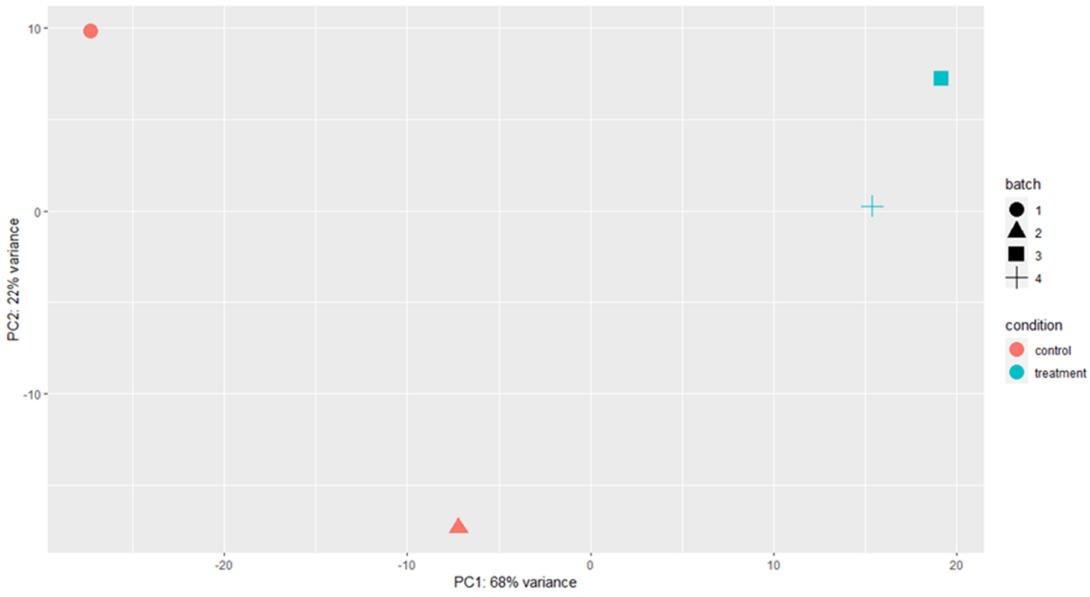


Figure 5.2 Principal component analysis (PCA) of the Brachyury knock-down experiment on 18hpf *P. lividus* samples. Different colors represent the experimental condition, while different shapes represent different batches.

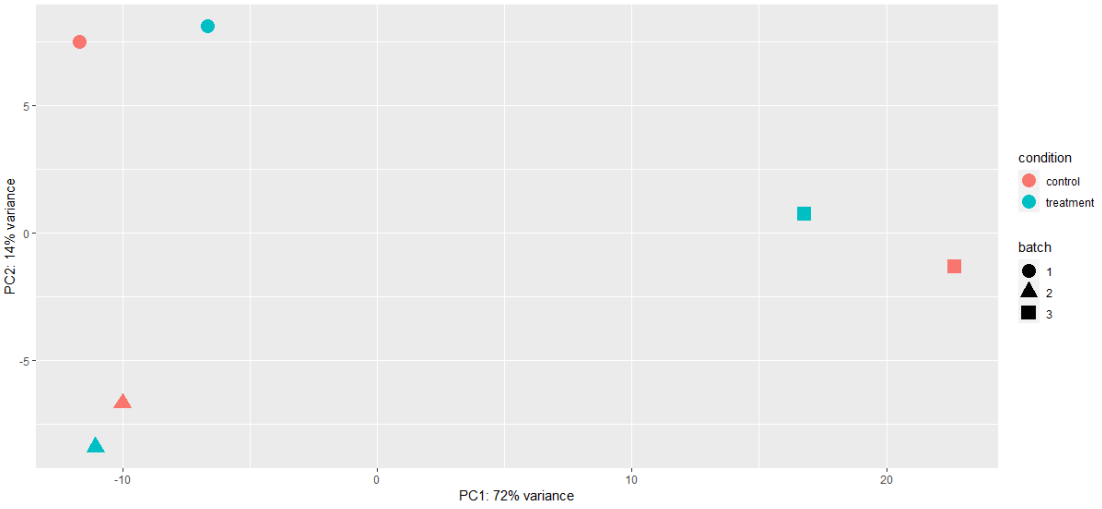


Figure 5.3 Principal component analysis (PCA) of the Brachyury knock-down experiment on 24hpf *P. lividus* samples. Different colors represent the experimental condition, while different shapes represent different batches PCA plots show close clustering of samples that come from the same batches of embryos.

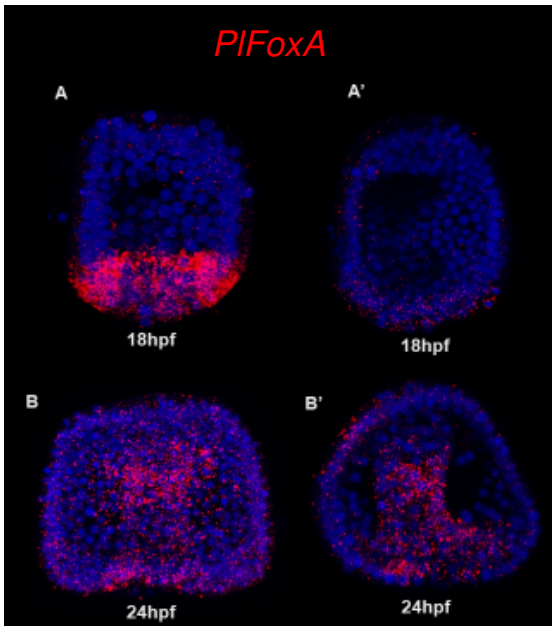


Figure 5.4 Whole-mount fluorescent *in situ* hybridization of *FoxA* at 18hpf and 24hpf in normal and Brachyury perturbed condition in *P. lividus*. A) and B) wild type condition; A') and B') Brachyury morpholino injected embryos. A) *FoxA* is normally expressed in the vegetal plate (endoderm). A') after Brachyury knock-down the expression of *FoxA* is reduced. B) oral ectoderm and gut expression of *FoxA* transcripts in wild type embryo, oral view. B') *FoxA* transcripts in the oral ectoderm and gut, lateral view. Courtesy of Periklis Paganos

5.2.2 Gene ontology analysis of Brachyury knock-down targets in P. lividus *at 18hpf and 24hpf*

To investigate Brachyury's role during the gastrulation of *P. lividus*, the biological processes affected by its perturbation were studied by performing Gene ontology enrichment analysis. Moreover, in an attempt to reconstruct the GRNs around this transcription factor, the differentially expressed genes dataset was filtered for transcription factors and signaling molecules using the annotation Build 8, that was used to annotate *S. purpuratus* (Kudtarkar and Cameron, 2017). In order to functionally annotate *P. lividus* gene IDs for both gene ontology enrichment analysis and functional annotation, they were converted to their corresponding *S. purpuratus*

gene ortholog names by blasting the *P. lividus* proteome to the proteome coming from *S. purpuratus* (Chapter 2).

In order to see which protein classes and biological processes were affected by Brachyury perturbation, gene ontology enrichment analysis was performed. Gene Ontology terms for *S. purpuratus* were used for the enrichment analysis in *PANTHER* (Mi *et al.*, 2016, 2019; Kudtarkar and Cameron, 2017).

At early gastrula stage (18hpf) 832 genes were affected by knocking-down Brachyury protein, while at the late gastrula (24hpf) the number was 15. At 18hpf 601 genes were downregulated (72%) and 231 genes were upregulated (28%) after the morpholino treatment. At 24hpf 10 genes were downregulated (67%) and 5 genes were upregulated (33%) after knocking down Brachyury protein.

At the early gastrula stage (18hpf), 832 genes were affected by knocking-down Brachyury protein, while at the late gastrula (24hpf), the number was 15. At 18hpf, 601 genes were downregulated (72%), and 231 genes were upregulated (28%) after the MASO treatment. At 24hpf, 10 genes were downregulated (67%), and 5 genes were upregulated (33%) after knocking down Brachyury protein.

At the early gastrula stage, the most highly affected biological processes were related to cellular processes and metabolism (Figure 5.5), which is in concordance with the most highly affected processes by Brachyury knock-down in *S. purpuratus* (Chapter 4). The cellular processes were terms associated with response to a stimulus, cellular component organization, and signal transduction. The GO terms related to metabolism were mostly part of organic substrate metabolic processes, cellular metabolic processes, primary metabolic processes, catabolic processes, biosynthetic processes, and small molecule metabolic processes. This suggests Brachyury's involvement in regulating the genes that maintain the embryo's energy supply and regulate normal growth (Marsh, Leong and Manahan, 1999). This could be a conserved role of Brachyury in regard to *S. purpuratus* (Chapter 4).

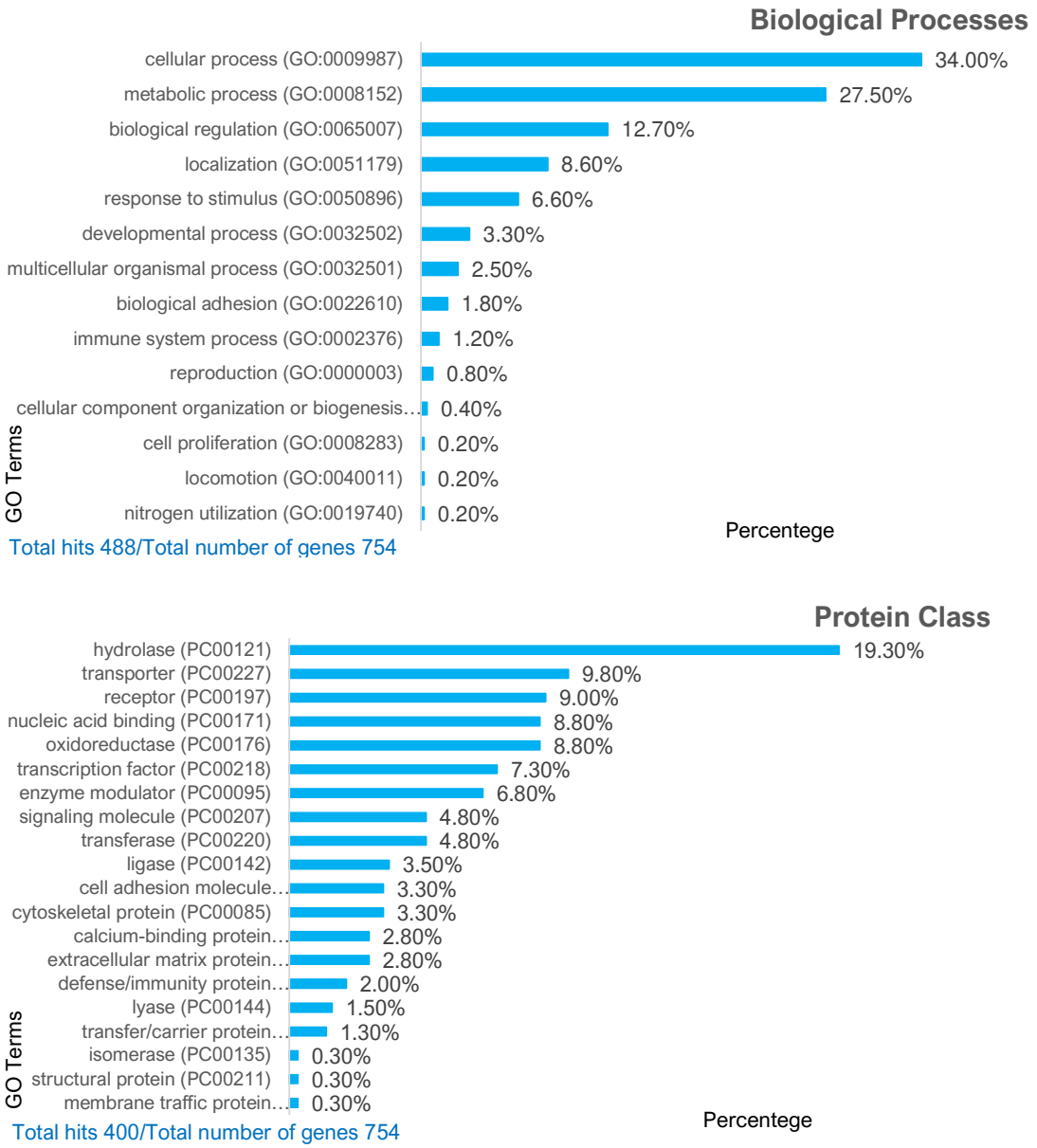


Figure 5.5 Gene Ontology terms analysis for 18hpf *P. lividus* Brachyury knock-down embryos. Each graph represents a specific category and contains specific GO terms and the percentage of genes associated with them. Lower left corner of each graphs shows the number of analyzed genes compared to the total number of differentially expressed genes. The upper graph represents the Biological Processes and the lower graph represents Protein class affected by the absence of Brachyury.

The top 10 protein classes affected by the Brachyury perturbation at 18hpf were hydrolases, transporters, receptors, nucleic acid binding proteins, oxidoreductases, transcription factors, enzyme modulators, signaling molecules, transferases, and ligases, which shows that Brachyury is an upstream regulator of multiple processes involved in transcriptional regulation and signaling pathways during development, and, as already described, metabolism (Figure 5.5).

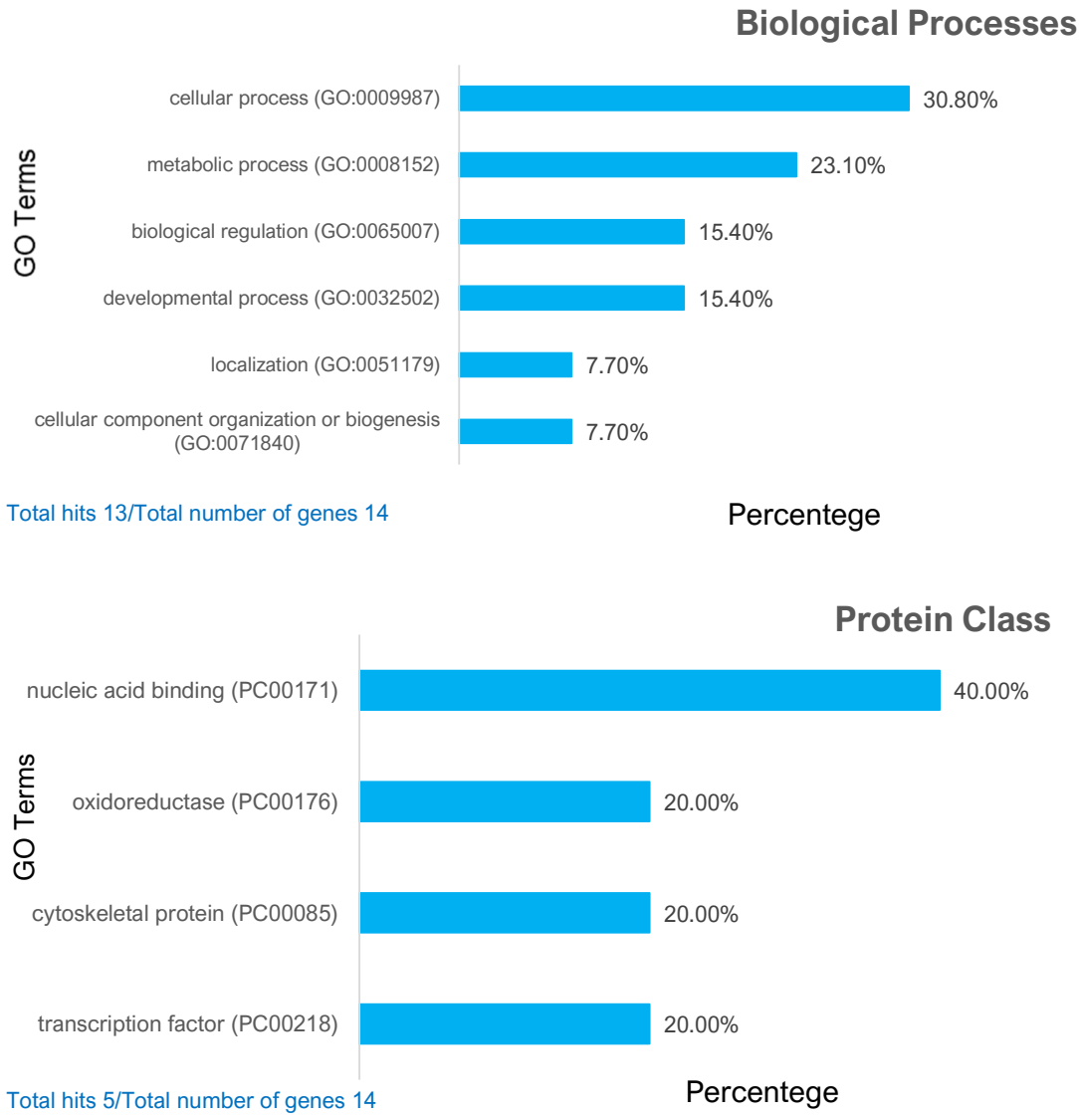


Figure 5.6 Gene Ontology terms analysis for 24hpf *P. lividus* Brachyury knock-down embryos. Each graph represents a specific category and contains specific GO terms and the percentage of genes associated with them. Lower left corner of each graphs shows the number of analyzed genes compared to the total number of differentially expressed genes. The upper graph represents the Biological Processes and the lower graph represents Protein class affected by the absence of Brachyury.

At the late gastrula stage (24hpf), few genes were affected by perturbation of Brachyury. Biological processes affected were related to cellular processes, metabolic processes, biological regulation, developmental processes, localization, and cellular component organization or biogenesis (Figure 5.6). The affected protein classes were nucleic acid binding, oxidoreductases, cytoskeletal proteins, and transcription factors (Figure 5.6). However, the small number of differentially expressed genes does not give enough information about the Brachyury involvement in regulating genes in the late gastrula. It is possible that the MASO effect was not so evident at the later stage and that many embryos recovered their normal development, as seen in Figure 5.3B'. In addition, it could be that the subject effects (the genomic make-up of each batch of embryos used as a biological replicate) are higher than the treatment effect.

5.2.3 Differentially expressed transcription factors and signaling molecules affected at 18hpf and 24hpf **P. lividus** *embryos after Brachyury perturbation*

In an attempt to reconstruct the GRNs downstream of Brachyury in *P. lividus* involved in developmental processes in the oral ectoderm and the posterior endoderm, the datasets were filtered for transcription factors and signaling molecules. At 18hpf, there were 98 TFs/signaling molecules affected by the Brachyury perturbation, out of which 79.5 % (78 genes) were downregulated and 20.5% (20 genes) were upregulated (Table 5.1). The expression patterns of 40 detected TFs/signaling molecules were described from the published literature.

Table 5.1 Differentially expressed transcription factors and signaling molecules after Brachyury knock-down at 18hpf in *P. lividus*. Each gene is described by its Pliv ID number, *S. purpuratus* equivalent gene name, logarithmic fold change (log2FC), spatial and temporal expression pattern and a reference. The list of perturbed genes is sorted based on the intensity of perturbation starting with the highest downregulation (the darkest blue color) and ending with the highest upregulation (the darkest red color).

Pliv_id	Gene (*S. purpuratus* equivalent)	Gene Description	log2FC	Spatial Expression	Reference
Pliv24776	Sp-Hh	hedgehog (hh) homolog	-7.177	Blastula: Veg2 endoderm; Gastrula: gut	Walton *et al.*, 2006; Peter and Davidson, 2011
Pliv29359	Sp-Nk7	NK7 homeobox	-6.945	Blastula and gastrula: PMC	Rafiq *et al.*, 2014; Dylus *et al.*, 2016
Pliv06407	SPU_021254	Sp-Egf/Lnb/Tm7/Gpcr_2	-6.769	unknown	n.a.
Pliv16127	Sp-ScratchX	scratch subfamily member X, z191	-6.408	Gastrula: tip of the archenteron	Materna, Swartz and Smith, 2013
Pliv00310	SPU_006724		-5.904	unknown	n.a.
Pliv14832	Sp-Ptf1a	pancreas specific transcription factor 1a-like, Sp-Ptf	-5.362	Late gastrula: midgut, apical organ and in scattered cells of the ciliary band	Perillo *et al.*, 2016
Pliv27874	SPU_016959		-5.198	unknown	n.a.
Pliv29568	Sp-Msx	msh-like homeobox, SpMsx	-5.134	Aboral ectoderm	Ben-Tabou de-Leon *et al.*, 2013
Pliv24655	Sp-Notchl6	Notch ligand-like 6, Notch-ligand6, Sp-Notchl6	-5.073	unknown	n.a.
Pliv23456	Sp-Apobec1	Apolipoprotein B mRNA editing enzyme	-4.129	Blastula: vegetal plate (veg2 meso and veg2 endo), Gastrula: gut, SMC	Peter and Davidson, 2011; Materna and Davidson, 2012
Pliv01006	Sp-Egf/Ig/Lnb/Tm7/Gpcr	Egf/Ig/Lnb/Tm7/Gpcr	-3.975	unknown	n.a.
Pliv08090	Sp-Nr2e6	nuclear receptor subfamily 2 group E-like	-3.837	unknown	n.a.

Pliv16468	Sp-Thytrprhr	Thyrotropin-releasing Hormone Receptor	-3.824	Gastrula: ciliary band	Arnone, unpublished; Wood *et al.*, 2018
Pliv16053	Sp-IrxA	iroquois homeobox A	-3.770	Blastula: aboral ectoderm, veg1 ecto, Early glastula: oral ectododerm	Howard-Ashby *et al.*, 2006; Su *et al.*, 2009; Ben-Tabou de-Leon *et al.*, 2013
Pliv29227	SPU_019621	Sp-Unc44_241	-3.739	unknown	n.a.
Pliv05386	SPU_027332	Sp-Clect/7Tm	-3.728	unknown	n.a.
Pliv03824	Sp-Hlx	H2.0-like homeobox	-3.604	unknown	n.a.
Pliv07861	Sp-Rxr	retinoid x receptor, NR2B	-3.518	unknown	n.a.
Pliv07163	Sp-Kr/Fa58C/Clect/Egf/Gpcr	Sp-Kr/Fa58C/Clect/Egf/Gpcr	-3.173	unknown	n.a.
Pliv20090	Sp-Klf7	Kruppel-like factor 7-like, Klf6/7, z86	-3.163	Late blastula-Early gastrula: Aboral and oral ectoderm	Materna *et al.*, 2006; Chen, Luo and Su, 2011
Pliv11409	Sp-Alx4	aristaless-like homeobox 4-like, Alx1/3/4-like	-3.069	PMC and SMC at the tip of the archenteron	Rafiq *et al.*, 2014; Khor and Ettensohn, 2017
Pliv10263	Sp-E78a	ecdysone-induced protein 78C (Drosophila)-like (partial), NR1E1-like	-3.049	unknown	n.a.
Pliv17036	Sp-Tll	tailless homolog, Nr2e1-like	-3.040	unknown	n.a.
Pliv20176	SPU_012160	Sp-Gpr21L_7, G protein-coupled receptor 21-like-7	-2.930	unknown	n.a.
Pliv26544	Sp-Hnf1aL	HNF1 homeobox (partial), hepatocyte nuclear factor 1-like, Sp-Hnf1aL, Sp-Hnf1L	-2.897	Late blastula and Early gastrula: veg1 endoderm Gastrula: midgut and hindgut	Howard-Ashby *et al.*, 2006; Peter and Davidson, 2011; Perillo *et al.*, 2016
Pliv19869	SPU_012714	Sp-Hypp_2046, hypothetical protein-2046	-2.684	unknown	n.a.
Pliv12306	Sp-Tgif	TGFB-induced factor homeobox	-2.623	Blastula: small micromeres; Late blastula and Early gastrula: veg2 endo; Late gastrula: tip of the archenteron, midgut and hindgut	Howard-Ashby *et al.*, 2006; Peter and Davidson, 2011
Pliv04872	Sp-Z56	zinc finger protein 56	-2.550	unknown	n.a.

Pliv25951	SPU_011559	Sp-Hypp_2046, hypothetical protein-2046	-2.538	unknown	n.a.
Pliv27356	Sp-Nr1m3	nuclear receptor subfamily 1 group M member 3, NR1-like	-2.529	unknown	n.a.
Pliv23891	Sp-Notchl3	Notch ligand-like 3, Notch-ligand3, Sp-Notchl3	-2.522	unknown	n.a.
Pliv19145	Sp-Glis1_1	PR domain containing 14-like, Sp-Z111, Sp-Prdm14	-2.472	unknown	n.a.
Pliv09718	SPU_011655	Sp-Lnb/7Tm/GpcrA	-2.405	unknown	n.a.
Pliv22990	Sp-Hypp_260	hypothetical protein-260	-2.390	unknown	
Pliv13648	Sp-Runx1t1	runt-related transcription factor 1 translocated to 1 (cyclin D-related) (human)-like, Cbfa2t1/2/3-like, ETO-like, nervy (nvy) homolog, Sp-Runx1t1	-2.381	unknown	n.a.
Pliv24136	Sp-Dlx	distal-less homeobox, Dll	-2.295	Blastula: veg1 ecto, aboral ecto; Gastrula: possible weak detection	Howard-Ashby *et al.*, 2006
Pliv02059	Sp-Fgf9/16/20	fibroblast growth factor, Sp-Fgf	-2.274	Late blastula and Early gastrula: Oral ectoderm	Röttinger *et al.*, 2008; Andrikou *et al.*, 2015
Pliv13690	Sp-Atbf1	AT-binding transcription factor 1-like, zinc finger homeobox, Sp-Zfhx, z30	-2.155	Oral and aboral ectoderm	Howard-Ashby *et al.*, 2006; Saudemont *et al.*, 2010
Pliv19949	Sp-Osr	odd-skipped-related, z121	-2.146	Mid-gastrula: midgut-hindgut boundary and hindgut	Materna *et al.*, 2006
Pliv24558	Sp-Hox5	homeobox 5	-2.120	unknown	n.a.

Pliv03782	Sp-FoxA	forkhead box A	-2.110	Blastula: oral ectoderm, vegetal plate, Gastrula: gut, stomodeum	Oliveri *et al.*, 2006; Tu *et al.*, 2006
Pliv29186	Sp-Nk1	NK1 homeobox	-2.108	Blastula: veg1 ecto (oral ectoderm border with endoderm); Gastrula: blastopore/vegetal plate	Minokawa *et al.*, 2004
Pliv03154	Sp-Nk2-2	NK2 homeobox 2, Sp-Nk2-2, Sp-NK2.2	-2.046	Late blastula: oral and aboral ectoderm	Howard-Ashby *et al.*, 2006; Chen, Luo and Su, 2011
Pliv09677	Sp-Scl	stem cell protein-like	-1.955	Blastula: veg2 meso	Solek *et al.*, 2013
Pliv12054	Sp-Myc	myelocytomatosis viral related oncogene, Myc/Mycn/Mycl-like	-1.943	Late blastula, Early gastrula: Veg2 endo; Gastrula: midgut	(Cui *et al.*, 2014)(Peter and Davidson, 2011a)
Pliv04959	Sp-Sfrp3/4l	Sp-Sfrp3/4l	-1.867	Late blastula nad Early gastrula: Vegetal plate (enododerm) and apical ectoderm; Late gastrula: apical ectoderm and endoderm	Illies *et al.*, 2002
Pliv17751	Sp-Grm3	Glutamate receptor, metabotropic 3, precursor	-1.863	unknown	n.a.
Pliv08080	Sp-Lmo4	LIM domain only 4	-1.820	unknown	n.a.
Pliv19216	Sp-Tbr	T-box brain-like, Eomes-like, Tbx21-like, ske-T	-1.819	Blastula and gastrula: PMC	Sharma and Ettensohn, 2010
Pliv03811	Sp-Six3	SIX homeobox 3, sine oculis-related homeobox 3/6, Sp-Six3/6	-1.793	Late blastula: endomesoderm and apical ectoderm; Late gastrula: tip of the archenteron, midgut foregut boundary and apical ectoderm	Poustka *et al.*, 2007; Wei *et al.*, 2009

Pliv26539	Sp-HgfL	hepatocyte growth factor-like, plasminogen-like	-1.785	unknown	n.a.
Pliv18858	SPU_012714	Sp-Hypp_2046, hypothetical protein-2046	-1.767	unknown	n.a.
Pliv25410	Sp-Wnt5	wingless-type MMTV integration site family member 5	-1.739	Late blastula: veg1 endo and veg1 ecto; Early and late gastrula: oral and aboral veg1 ectoderm	Mcintyre *et al.*, 2013; Cui *et al.*, 2014
Pliv17221	Sp-Hypp_2098	hypothetical protein-2098	-1.737	unknown	n.a.
Pliv14169	SPU_008921		-1.728	unknown	n.a.
Pliv10300	Sp-Ese	epithelium-specific ets factor-like	-1.701	Blastula: veg2 meso; Glastula: SMC	Rizzo *et al.*, 2006
Pliv04889	Sp-Gsc	goosecoid homeobox	-1.666	Oral ectoderm	Angerer *et al.*, 2001
Pliv27661	Sp-Znf608L	zinc finger protein 608-like; KIAA1281 protein-like	-1.652	unknown	n.a.
Pliv05487	Sp-FoxN2/3	forkhead box N2/3, checkpoint suppressor 1	-1.648	Up to mesenchyme blastula: mesoderm; Late blastula and Early gastrula: veg2 endo; late gastrula: gut	Tu *et al.*, 2006; Rho and McClay, 2011
Pliv23016	Sp-Hnf6	onecut homeobox, Onecut1/2/3-like	-1.637	Gastrula: ciliary band	Otim *et al.*, 2004
Pliv29089	Sp-Hhex	hematopoietically expressed homeobox, Sp-HEX, Hhex, Prh, Hmph	-1.608	Blastula: SMC; Gastrula: tip of the archenteron)	Howard-Ashby *et al.*, 2006; Poustka *et al.*, 2007
Pliv24880	Sp-Plek	pleckstrin	-1.552	unknown	n.a.
Pliv13082	SPU_019991	Sp-Unc44_141, ankyrin2,3/unc44-like-141	-1.532	unknown	n.a.
Pliv09905	Sp-Hmha1	Minor histocompatibility antigen HA1, HLA-HA1	-1.532	unknown	n.a.

Pliv02950	Sp-Rgs7	regulator of G-protein signaling 7, RGS-7	-1.503	unknown	n.a.
Pliv30061	SPU_012714		-1.498	unknown	
Pliv17449	Sp-Blimp1	B lymphocyte induced maturation protein-like, PRDM1-like, blimp1/krox, SpKrox1	-1.464	Late blastula and early gastrula: veg2 endoderm; Late gastrula: hindgut	Livi and Davidson, 2007; Smith *et al.*, 2008; Peter and Davidson, 2011
Pliv30030	Sp-Kndc1	kinase non-catalytic C-lobe domain (KIND) containing 1	-1.423	unknown	n.a.
Pliv13715	Sp-Rreb1	ras responsive element binding protein 1-like, z48, pebbled homolog	-1.382	Late blastula: mesoderm	(Materna *et al.*, 2006)
Pliv03464	Sp-Wrch	Ras like protein	-1.281	unknown	n.a.
Pliv12703	Sp-Gcm	glial cells missing homolog, Gcm1/2-like	-1.274	Blastula: veg2 meso; Gastrula: SMC	Ransick *et al.*, 2002
Pliv24397	Sp-GataC	GATA binding protein C	-1.228	Blastula: SMC; Glastula: SMC, PMC	Davidson *et al.*, 2002; Solek *et al.*, 2013
Pliv21163	Sp-Spry/Socscp	Sprouty homolog	-1.215	unknown	n.a.
Pliv24364	Sp-Elk	Elk1/3/4-like	-1.184	Blastula: vegetal plate (most probably only veg2 meso); gastrula: ubiq	Rizzo *et al.*, 2006
Pliv06984	Sp-Irsp53/58	Insulin receptor substrate p53/p58-like, brain-specific angiogenesis inhibitor 1-associated protein 2-like, Baiap2/2L1/2L2, FLAF3-like, Sp-Irsp53/58	-1.142	unknown	n.a.
Pliv03381	Sp-Ets1/2	v-ets erythroblastosis virus E26 oncogene homolog 1/2-like	-1.105	Blastula: veg2 meso, PMC; Glastula: SMC, PMC	Rizzo *et al.*, 2006

Pliv19254	Sp-Frizz5/8	rizzled homolog 5 and 8-like, Sp-Frizz5/8, SpFz5/8	-1.097	Blastula: Veg2 meso, apical ectoderm; Late gastrula: apical and oral ectoderm	Cui *et al.*, 2014; Arnone, unpublished
Pliv25286	Sp-Phb1	paired homeodomain 1	1.190	unknown	n.a.
Pliv27980	Sp-Trip4	thyroid hormone receptor interactor 4	1.359	unknown	n.a.
Pliv14337	Sp-Znf622	zinc finger protein 622	1.835	unknown	n.a.
Pliv14450	Sp-Nfxl1-2	nuclear transcription factor, X-box binding-like 1-2	2.043	unknown	n.a.
Pliv11883	Sp-Z65	zinc finger protein 65	2.052	ubiq	Materna *et al.*, 2006
Pliv04534	Sp-Znf259l	zinc finger protein 259	2.055	unknown	n.a.
Pliv14248	Sp-Nfxl1-2	nuclear transcription factor, X-box binding-like 1-2	2.075	unknown	n.a.
Pliv07255	Sp-Hypp_1134		2.101		n.a.
Pliv16485	Sp-Cebpa	CCAAT/enhancer binding protein alpha-like, C/EBP alpha-like, Cebpa/b/d-like	2.197	unknown	n.a.
Pliv04328	Sp-Kiaa1822	hedgehog interacting protein-like, Hhipl1/2-like, Sp-Hhipl	2.202	Most probably same expression as Hh	Walton *et al.*, 2006; Peter and Davidson, 2011
Pliv15631	SPU_011359	Sp-Unc44L_72, ankyrin2,3/unc44-like-72	2.377	unknown	n.a.
Pliv22976	Sp-Grm3	Glutamate receptor, metabotropic 3, precursor	2.874	unknown	n.a.
Pliv00801	Sp-Ap4	transcription factor AP-4, Tfap4, activating enhancer binding protein 4, cropped homolog, Sp-Tcfap4L	3.751	unknown	n.a.
Pliv12716	SPU_023629	Sp-Rxfp1_2, relaxin/insulin-like family peptide receptor 1-2	4.080	unknown	n.a.

Pliv06716	Sp-Drd4L	dopamine receptor D4-like-1	4.536	unknown	n.a.
Pliv27116	SPU_002094	Sp-CckarL, cholecystokinin A receptor-like	4.848	unknown	n.a.
Pliv08396	SPU_023590	Sp-Rej4_1, receptor for egg jelly 4	4.983	unknown	n.a.
Pliv07841	Sp-FtzF	nuclear receptor subfamily 5 group A, Ftz-F1 homolog, Sp-FtzF	5.018	unknown	n.a.
Pliv20784	Sp-Mib2_1	mindbomb-like with protein kinase domain, Mib1/2-like	5.336	Most probably Notch signaling pathway	n.a.
Pliv21436	Sp-Mib2_1	mindbomb-like with protein kinase domain, Mib1/2-like	6.910	Most probably Notch signaling pathway	n.a.

Out of those 40, 1 gene was found to be ubiquitously expressed, 15 were found to be expressed in the ectoderm, 14 in the endoderm, and 16 in the mesoderm of the sea urchin embryo (Table 5.1). This dataset indicates that Brachyury could act as an activator of endodermal, mesodermal, and ectodermal genes, respectively. Nonetheless, it should be taken into account that Brachyury is not expressed in any mesodermal cell types making the direct function as an activator in mesodermal cells unlikely. Moreover, since the replicates for this differential expression analysis were selected “manually,” it could be possible that many of the detected perturbed genes were, in fact, not really perturbed. By removing the replicates that did not have *FoxA* downregulated, PCA analysis showed higher variability of the untreated vs. treated samples. This could be explained not just by the differences in the effect of perturbation, but there could be the errors that appeared after selecting replicates with bias.

Table 5.2 Differentially expressed transcription factors and signaling molecules after Brachyury knock-down at 24hpf in *P. lividus.* Each gene is described by its Pliv ID number, *S. purpuratus* equivalent gene name, logarithmic fold change (log2FC), spatial and temporal expression pattern and a reference. The list of perturbed genes is sorted based on the intensity of perturbation starting with the highest downregulation (the darkest blue color) and ending with the highest upregulation (the darkest red color).

Pliv_id	Gene (*S. purpuratus* equivalent)	Gene Description	log2FC	Spatial Expression	Reference
Pliv17335	Sp-Bra	Brachyury	1.269	Blastula: veg endo, oral ectoderm; Glastula: blastopore, hindgut, stomodeum	Croce, Lhomond and Gache, 2001; Gross and McClay, 2001; Rast *et al.*, 2002; Andrikou, 2012
Pliv28511	Sp-FoxY	forkhead box Y, forkhead C-like	0.780	Blastula: small micromeres, SMC; gastrula: SMC, tip of the archenteron	Materna, Swartz and Smith, 2013; Annunziata *et al.*, 2014; Andrikou *et al.*, 2015

Regarding 24hpf knockdown embryos, only two genes were found, whose expression pattern has been known. One is *the Bra* gene itself, and the other one is *FoxY* expressed in the secondary mesenchyme cells at the tip of the archenteron. Both of those genes were upregulated.

5.3 Conclusions

This chapter provided some significant hints to the possible role of Brachyury in the sea urchin *P. lividus* that can be the basis of future projects regarding this question. This dataset is preliminary, and more experiments are needed to safely draw conclusions and compare the role of Brachyury in the two sea urchin subject species of this study.

The main principle of the cross-species comparison approach I chose to perform is based on the reconstruction of the individual gene regulatory networks after perturbation of gene expression, in this case, knockdown of Brachyury, and comparison of those GRNs. Due to the different responses of embryos to the same MASO dosage, the variability of phenotypes, and the strong batch effect between

wild type and injected embryos, the differential gene expression analysis could not be performed in the same unbiased way as in *S. purpuratus* (Chapter 4). For all those reasons, the reconstruction of a solid and reliable GRN is not possible, at least until the number of replicates increases, and the phenotypic plasticity of the embryos is addressed. Despite this, the results of the analysis and the GRNs for the 18hpf embryo that were reconstructed based on it are presented here (Figure 5.7), as an example of how we could use the cross-species GRN comparison in untangling the evolution of the role of a gene in evolution at the microscale.

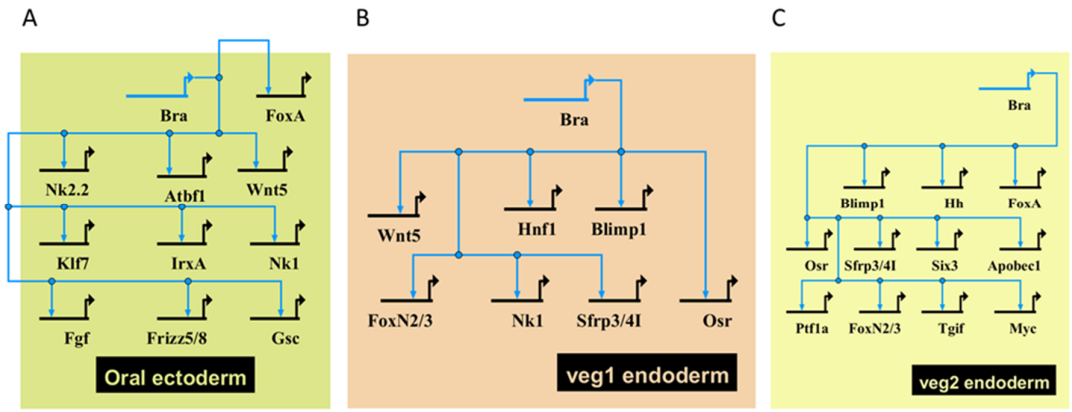


Figure 5.7 Schematic representation of the putative oral ectoderm (A) and vegetal plate (B & C) Gene regulatory networks around Brachyury at 18hpf in *P. lividus* using BioTapestry. (http://www.biotapestry.org/). Inputs coming from the same gene are shown. The arrows represent positive inputs.

On the other hand, if future experiments validate this dataset, it could mean that the Brachyury-centered oral ectoderm and vegetal plate GRNs got rewired and that the role of Brachyury changed through echinoderm evolution. This change could be reflected in turning Brachyury mostly into an activator of gene expression and an activator of mesodermal genes, and thus mesodermal cell fate, which is the opposite of what it seems to be the case in *S. purpuratus* (Chapter 4). At this moment, however, no firm conclusions could be made. The presented results should be taken with a grain of salt because the *P. lividus* samples have high batch variability. In the future, this problem could be overcome by introducing a higher number of biological

replicates or avoiding the use of wild animals but instead using the animals bred in the lab that show similar genetic makeup.

Contribution statement

Dr. Maria I. Arnone performed the microinjections. Perkilis Paganos, a PhD student from Arnone's lab, took care of the embryo cultures, imaged the phenotypic differences, and collected RNA for sequencing. P. Paganos performed the whole-mount *in situ* hybridization experiment. The author of this thesis performed a literature search, RNA-seq analysis, and drafted the GRNs.

CHAPTER 6

DISCUSSION

This chapter contains the discussion of the results obtained, including the issues identified and the possible explanations for differences in the Brachyury effect following morpholino microinjections. Moreover, this chapter states future research directions and draws overall conclusions.

6.1 Systems biology approach identifies direct Brachyury targets

The rapid development of a systematic point of view in biology started at the beginning of the 21st century, although its roots were settled much before, with the discovery of the chromatin and the structure of the DNA molecule, as well as the development of the methods of gene manipulation (Arkin and Schaffer, 2011). Before these advances and the advent of sequencing technologies, research in biological sciences focused on individual genes, proteins, metabolites, organs, etc. Each “component” was studied one at a time – this approach was slow, laborious, and biased. There is no single gene, a product of a gene, or a process in an organism that acts alone. Studying the complex interplay between different system components concerns systems biology, a holistic discipline, which connects individual components and requires complex tools. Omics can be described as various experimental approaches used to investigate the roles, interconnections, and actions of different molecules that compose an organism. These approaches include high-throughput methods used for investigating the functions of genes - genomics, genome-wide epigenetic modifications of DNA – epigenomics, mRNAs - transcriptomics, proteins - proteomics, and metabolites - metabolomics (Ram, Mendelsohn and Mills, 2012; Tavassoly, Goldfarb and Iyengar, 2018).

This integrative, or systems biology approach is the main method of my research project. It combines transcriptomics or RNA-seq (Kukurba and Montgomery, 2015),

Assay for Transposase-Accessible Chromatin using sequencing, or ATAC-seq (Buenrostro *et al.*, 2015), and Chromatin immunoprecipitation sequencing or ChIP-seq (Barski *et al.*, 2007). Each method was described in more detail in different chapters throughout this thesis.

The approach used has several major steps and datasets obtained at each step, starting from the transcription factor. The MASO injection is the starting point of such an approach. This is followed by a simple phenotype assessment. The phenotype differences between the wild type and treated samples confirm that the perturbed protein does have a function during development. Then, this is followed by differential RNA-seq experiments to detect in-bulk what genes are affected directly or indirectly. ChIP-seq for the transcription factor of interest can be performed to identify where the protein is actually bound to the DNA, then ATAC-seq, which shows open chromatin, frequently associated with CRMs, shows which of the ChIP-seq locations are likely to be real and really be TF-associated CRMs, telling, which of the bulk targets are direct. Then localization experiments, such as in situ hybridization or immunohistochemistry, can show the cells and the tissues, in which the targets identified by omics methods after morpholino injection are expressed.

This approach looks at many aspects of the role of Brachyury since it concerns both its protein, which was discussed in Chapter 3 in terms of stability and interaction with other proteins and, in Chapter 4 in terms of transcription factor motifs recognized by the Brachyury transcription factor, which also can be traced back to the structure of the protein. It also concerns mRNA in terms of the affected transcripts identified by RNA-seq analysis and also in part by in situ hybridization experiments. Finally, this approach also concerns the DNA by identifying CRMs using the ChIP-seq/ATAC-seq combined data set and, in a way, through the identification of affected genes, which can be deduced by looking at the mRNA of these genes affected by perturbation.

Therefore, such an approach covers a vast number of aspects of the transcription factor role: DNA, RNA, protein, tissue, organism – which all interplay giving information about this transcription factor in terms of systems biology. A comparison

of these aspects gives evolutionary insight into the transcription factors' function and the GRN wiring, as mentioned in Chapters 4 and 5.

6.2 Brachyury is an activator of endodermal and ectodermal fates, but repressor of mesodermal fate in the sea urchin

Using morpholino oligonucleotide to prevent *SpBra* translation has shed light on this protein's role during sea urchin development. The germ layers from which the affected genes are derived suggest that, at least, in *S. purpuratus*, Brachyury is an activator of ectodermal and endodermal lineage genes, leading to specification of the tissues derived from these germ layers. At the same time, mesodermal genes showed that SpBra is a repressor of mesoderm specifying genes.

The GRNs drafted for these germ layers at different time points indicate that Bra could be one of the high-level actors in germ layer specification since it controls multiple transcription factors and signaling pathway components, as shown by the GRN drafted from RNA-seq differential analysis after Bra perturbation and combination with ChIP-seq/ATAC-seq analysis. These targets, in turn, by their nature, will have their own targets. This shows the key position of Brachyury in the regulation of cell fate specification.

6.2.1 Sea urchin Brachyury fits into a broad evolutionary scenario

Starting from its discovery in the developing mouse embryos, *brachyury* was considered a key mesodermal gene since the further description was consistent in vertebrates (human, mouse, frog, zebrafish) and non-vertebrate deuterostomes (urochordates and cephalochordates)(Wilkinson, Bhatt and Herrmann, 1990; Kispert *et al.*, 1994; Holland *et al.*, 1995; Conlon *et al.*, 1996; Corbo *et al.*, 1997; Messenger *et al.*, 2005). Ambulacrarians, composed of echinoderms and hemichordates, are crucial basal deuterostomes, and they show different expression patterns of *brachyury*, where this TF is primarily expressed in oral and anal regions. Later during

development, hemichordates show mesodermal expression (Peterson, Cameron, *et al.*, 1999).

What about the protostomes and non-bilaterians?

Both groups are connected with the shared expression pattern of *brachyury*. In cnidarians, one of the most important non-bilaterian groups, *brachyury* expression is in the ectoderm – endoderm boundary, precisely in the blastopore region. Ctenophores, currently with an enigmatic phylogenetic position, then mollusks, annelids, and arthropods show the conserved pattern of *brachyury* expression: it is always found in the blastopore (future mouth) and anal regions (Kispert *et al.*, 1994; Arendt, Technau and Wittbrodt, 2001; Lartillot *et al.*, 2002; Lengyel and Iwaki, 2002; Shinmyo *et al.*, 2006; Berns *et al.*, 2008; Yamada *et al.*, 2010; Arenas-Mena, 2013). Some protostome groups, such as insect order Diptera, seem to have co-opted *brachyury* in the mesoderm formation (Kusch and Reuter, 1999; Arendt, Technau and Wittbrodt, 2001). In all studied insect species, *brachyury* is quite conserved: it is found in the blastopore and the hindgut (Kispert *et al.*, 1994; Shinmyo *et al.*, 2006; Berns *et al.*, 2008). However, the fruit flies show an additional domain of expression in the caudal visceral mesoderm. Moreover, this specificity is also seen in some annelids. It could be possible that the inductive mesodermal function of *brachyury* was co-opted *de novo* multiple times during metazoan evolution – in annelids, dipterans, hemichordates, and the last common ancestor of chordates.

Brachyury *as a pan-blastoporal and a pan-ectodermal gene*

Why should we not refer to *brachyury* only as a “crucial mesodermal” gene anymore? Maybe the answer can connect what is shared between protostomes and deuterostomes. Based on the data gained in this thesis work, Brachyury's function shows the inductive function in the ectoderm and posterior endoderm formation, while the effect on mesodermal genes is repressing. In all metazoan groups, the expression of *brachyury* is always initiated in the blastopore region. Even in vertebrates, before it is expressed in the mesoderm, the expression is seen around the blastopore. Comparatively speaking, the vertebrate *brachyury* is expressed in the neuromesodermal precursors before they differentiate into separate germ layers. During vertebrate development, Brachyury serves as a neuroectodermal

repressor and mesodermal activator. A study by Gentsch et al. has shown that Brachyury depleted embryos fail to form mesodermal derivatives and form an oversized neural tube (Gentsch *et al.*, 2013). Therefore, Brachyury should be described as a pan-blastoporal transcription factor first, and then as pan-ectodermal and mesodermal.

6.3 Changes in amino acid content could lead to different protein stability

As identified previously by Dr. Andrikou, Brachyury protein is detectable in the developing archenteron of the sea urchin *S. purpuratus* gastrula, and this, however, was not observed for Brachyury from other sea urchin species studied, as stated in Chapter 3. Such observation could be due to the different stability of this transcription factor protein in different species.

Multiple amino acid substitutions were identified, many of which change hydrophobicity or hydrophilicity, charge, and size of the amino acid at a particular location in the protein's primary structure. The effect that the primary sequence has on the secondary structure was also discussed, and the differences in likely secondary structures between the species were also identified. Secondary structure affects the tertiary structure, the actual folding of the protein, affecting this structure's stability. This folding change could also lead to the different sensitivity of antibodies mentioned in Chapter 3 because changes like that could make certain parts of the protein more accessible and detectable by an antibody.

In addition, the evolutionary changes in the amino acid sequence observed between the different species suggest that detected substitutions can be targets for post-translational modifications or provide targets for degradation.

A number of experiments can be done to assess if these *in silico* predictions are correct in the future. The Brachyury protein's actual 3D structures can be obtained using X-ray crystallography (Ilari and Savino, 2008). As mentioned before, transgenesis using CRISPR-Cas9 can be used to express SpBra in *P. lividus*. The

stability of protein can also be measured *in vivo* in the future using the known techniques, for example, using FlAsH labeling (Ignatova and Gierasch, 2009) or or tandem fluorescent protein timers to assess protein turnover (Khmelinskii *et al.*, 2012).

6.4 Importance of Brachyury protein sequence evolution for protein-protein interactions

As described previously in Chapters 1 and 3, Brachyury-Smad1 interaction is one of the most important co-transcriptional mechanism in metazoan development. Interaction between Smad1, Smad2/3, Eomesodermin, and Brachyury has been described as a crucial mechanism determining mesoderm inductive capabilities of Brachyury and endodermal inductive capabilities of Eomesodermin in vertebrate development (Faial *et al.*, 2015). In addition, it has been shown that the ectodermal repressing action of Brachyury is mediated by Smad1, where Brachyury and Smad1 cooperate in activating the neural repressor TF *ventx1*.1 during *Xenopus* development (Shiv Kumar *et al.*, 2018).

The echinoderm Brachyury protein sequence comparison showed that all tested echinoderms have a conserved consensus sequence in their N-terminal domain, which is able to interact with Smad1 echinoderm homolog Smad1/5/8. It could be possible that Brachyury and Smad interacting proteins evolved together. This hypothesis can be supported by the observation made in *D. melanogaster* N-terminus of Brachyenteron that lacks the Smad1 interacting consensus sequence. Not surprisingly, as mentioned above, in dipterans, Brachyury gained an additional domain of expression (visceral mesoderm). Perhaps this additional Brachyury function is connected to the loss of Smad1 interaction. Moreover, one of the rare protostome groups that had lost Brachyury is the nematode group. In *C. elegans*, *brachyury's* function was probably co-opted by the *mab-9* (Woollard and Hodgkin, 2000). Even though some previous studies confirmed the presence of the Smad1 orthologue in *C. elegans*, due to similar naming (*sma1*), there was confusion (Das,

Maduzia and Padgett, 1999; Massagué, Seoane and Wotton, 2005). When the author of this thesis blasted MH2 domains containing the consensus Smad1 interacting amino acid residues from various species against the *C. elegans* proteome, there were no significant hits to point to a Smad1 *C. elegans* orthologue (data not shown). It seems that *C. elegans* had lost both Brachyury and Smad1. This finding might point out that Brachyury-Smad1 interaction is one of the most important co-transcriptional mechanisms guiding animal development.

This protein-protein interaction should be considered for further exploration in future Brachyury studies.

6.5 Possible explanations to differences in MASO response between the two species

The PCA plots described in Chapter 5 suggest that the batches of *P. lividus* zygotes injected with the PlBra MASO do not respond to the MASO in the same way and that other factors contribute to most of the variability between control and treatment conditions (wild-type and Bra MASO injected). This could be due to imperfect MASO design because of the incomplete genome annotation, as sequence could be missing in the location immediately upstream of the Brachyury TSS to the region where the MASO was designed. That could lead to the morpholino having decreased activity compared to SpBra MASO, which was similarly designed, but the annotation of SpBra is of higher confidence.

The major difference between SpBra and PlBra targets shows that in *S. purpuratus*, Bra plays a mesodermal repressor role, while there is not much evidence for the same role in *P. lividus*, as it seems to be mostly an activator in this species. If the observed role differences of Bra protein in two species is due to biological species differences, this could be explained by Hedgehog (Hh) signaling. *Hh* is one of the top targets after Bra perturbation in *P. lividus,* and is known to affect mesoderm derived tissue fates (Walton *et al.*, 2009). Differences in mesodermal targets between the two species could be explained by Bra acquiring, in the case of *P.*

lividus, or losing in the case of *S. purpuratus, Hh* as a target. This could be due to CRM differences, e.g., due to mutations that add or remove the Brachyury binding site from the CRM controlling *Hh*. The Hh hypothesis could be assessed in the future.

Some genes affected by Bra perturbation in the early gastrula of *P. lividus* and the late gastrula of *S. purpuratus* could be an example of heterochrony, although its significance in the sea urchin evolution is unclear and should be addressed in the future.

6.6 Conclusions

This project has resulted in reconstructing the GRN downstream of Brachyury in *S. purpuratus* using various omics approaches, namely RNA-seq differential analysis, ChIP-seq, and ATAC-seq analyses, through a combinatorial approach. Therefore, direct and potential indirect targets of the *S. purpuratus* Brachyury in multiple tissues and cell types derived from all three germ layers were identified.

Furthermore, preliminary datasets for *P. lividus* were also produced, which allowed assembling initial drafts of GRNs downstream of *P. lividus* Brachyury. However, no definite conclusion can be drawn due to technical issues, spanning from the genome assembly to the actual knock-down experiment. Thus, the analyses presented in Chapter 5 are speculative, and further studies are needed to shed light on Brachyury's role in *P. lividus*.

Evolutionary comments on the obtained data between the two species and the published data from other animal species were made, stating the changes in Brachyury's mesodermal role. Moreover, this work has shown that Brachyury's ancestral metazoan function is in regulating morphogenetic movements during gastrulation and in the formation of mouth and anus.

Multiple points of view and aspects of the Brachyury role looking at protein, RNA, DNA, and cell type level constitutes the used systems biology approach. Lastly, this project identified directions for future research.

NON-BOOK COMPONENT

The non-book component is provided on the USB drive and it contains the detailed results described throughout Chapters 3, 4, and 5:

1. Pairwise alignment of *P. lividus* and *S. purpuratus* Brachyury protein sequence.

   S_purpuratus_P_lividus_pairwise_alignment.txt

2. Differential RNA-seq full results tables (*P. lividus* 18hpf, *P. lividus* 24hpf, *P. lividus* standard control MASO 24hpf, *S. purpuratus* 27hpf, *S. purpuratus* 48hpf, *S. purpuratus* standard control MASO 48hpf). Differential RNA-seq filtered and annotated result tables (*P. lividus* 18hpf, *P. lividus* 24hpf, *P. lividus* 24hpf standard control MASO, *S. purpuratus* 27hpf, *S. purpuratus* 48hpf, *S. purpuratus* 48hpf standard control MASO).

   ihw_pl18wt_bramo_results.csv

   ihw_pl24wt_bramo_results.csv

   IHW_pl24wt_ctrlmo_results.csv

   IHWsp27wt_bra_results.csv

   IHWsp48wt_bra_results.csv

   IHWsp48ctrlmo_results.csv

   pl18_annotated.xlsx

   pl24_annotated.xlsx

   IHW_pl24wt_vs_ctrlmomo_results_filtered_padj005.csv

   sp27bra_annotated.xlsx

   sp48bra_annotated.xlsx

   IHW_sp48ctrlmo_filtered_padj005.csv

3. Tables containing *S. purpuratus* ChIP-seq – ATAC-seq intersected peaks filtered by the differential RNA-seq results and presence of Brachyury motifs; annotated by gene name (24/27hpf and 48 hpf).

   chip_atac_seq_intersect_bra_motifs_annotated_peaks_24_27wt.xlsx

   chip_atac_seq_intersect_bra_motifs_annotated_peaks_48wt.xlsx

4. HTML report of *de novo* discovered motif from 24hpf Brachyury ChIP-seq high stringency dataset contained in the directory named *high_sig_24_dreme*.

5. Full echinoderm Brachyury protein alignment with marked important amino acid residues in *S. purpuratus* and *H. pulcherrimus* sequences

   Figure 3.6.1.png

PUBLICATIONS

Magri M.S., Voronov D., Ranđelović J., Cuomo C., Gómez-Skarmeta J.L., Arnone M.I. (2021) ATAC-Seq for Assaying Chromatin Accessibility Protocol Using Echinoderm Embryos. In: Carroll D.J., Stricker S.A. (eds) Developmental Biology of the Sea Urchin and Other Marine Invertebrates. Methods in Molecular Biology, vol 2219. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-0974-3_16

I have contributed to the adaptation of the protocol for sea urchin embryos and was involved in ATAC-seq library generation for the two echinoid species.

REFERENCES

Adell, T. *et al.* (2003) 'Isolation and characterization of two T-box genes from sponges, the phylogenetically oldest metazoan taxon', *Development Genes and Evolution*, 213(9), pp. 421–434. doi: 10.1007/s00427-003-0345-5.

Adli, M. and Bernstein, B. E. (2011) 'Whole-genome chromatin profiling from limited numbers of cells using nano-ChIP-seq', *Nature Protocols*. doi: 10.1038/nprot.2011.402.

Andreou, A. M. *et al.* (2007) 'TBX22 missense mutations found in patients with X-linked cleft palate affect DNA binding, sumoylation, and transcriptional repression', *American Journal of Human Genetics*, 81(4), pp. 700–712. doi: 10.1086/521033.

Andrews, S. and Babraham Bioinformatics (2010) 'FastQC: A quality control tool for high throughput sequence data', *Manual*. doi: citeulike-article-id:11583827.

Andrikou, C. (2012) *Evolution Of Mesoderm Specification And Myogenesis In The Sea Urchin Embryo* PhD Thesis The Open University.

Andrikou, C. *et al.* (2013) 'Myogenesis in the sea urchin embryo: The molecular fingerprint of the myoblast precursors', *EvoDevo*, 4(1). doi: 10.1186/2041-9139-4-33.

Andrikou, C. *et al.* (2015) 'Logics and properties of a genetic regulatory program that drives embryonic muscle development in an echinoderm', *eLife*. doi: 10.7554/eLife.07343.

Angerer, L. M. *et al.* (2001) 'Sea urchin goosecoid function links fate specification along the animal-vegetal and oral-aboral embryonic axes', *Development*.

Annunziata, R. (2011) *Evolution of a Gene Regulatory Network that Controls Gut Development, Differentiation And Functioning*. PhD Thesis The Open University.

Annunziata, R. *et al.* (2014) 'Pattern and process during sea urchin gut morphogenesis: The regulatory landscape', *Genesis*, 52(3), pp. 251–268. doi: 10.1002/dvg.22738.

Annunziata, R. and Arnone, M. I. (2014) ‘A dynamic regulatory network explains ParaHox gene control of gut patterning in the sea urchin’, *Development*, 141(12), pp. 2462–2472. doi: dev.105775 [pii]\r10.1242/dev.105775.

Arenas-Mena, C. (2013) ‘Brachyury, Tbx2/3 and sall expression during embryogenesis of the indirectly developing polychaete Hydroides elegans’, *International Journal of Developmental Biology*, 57(1), pp. 73–83. doi: 10.1387/ijdb.120056ca.

Arendt, D., Technau, U. and Wittbrodt, J. (2001) ‘Evolution of the bilaterian larval foregut.’, *Nature*, 409(6816), pp. 81–85. doi: 10.1038/35051075.

Arkin, A. P. and Schaffer, D. V. (2011) ‘Network news: Innovations in 21st century systems biology’, *Cell*. Elsevier Inc., 144(6), pp. 844–849. doi: 10.1016/j.cell.2011.03.008.

Arnone, M. I. *et al.* (2006) ‘Genetic organization and embryonic expression of the ParaHox genes in the sea urchin S. purpuratus: Insights into the relationship between clustering and colinearity’, *Developmental Biology*, 300(1), pp. 63–73. doi: 10.1016/j.ydbio.2006.07.037.

Arnone, M. I., Byrne, M. and Martinez, P. (2015) ‘Echinodermata’, in Wanninger, A. (ed.) *Evolutionary Developmental Biology of Invertebrates*. 6th edn. Vienna: Springer Vienna, pp. 1–58. doi: 10.1007/978-3-7091-1856-6_1.

Arnone, M. I., Dmochowski, I. J. and Gache, C. (2004) ‘Using reporter genes to study cis-regulatory elements.’, *Methods in cell biology*. United States, 74, pp. 621–652.

Artavanis-Tsakonas, S., Rand, M. D. and Lake, R. J. (1999) ‘Notch signaling: Cell fate control and signal integration in development’, *Science*. doi: 10.1126/science.284.5415.770.

Baer, K. E. von (1828) *Über Entwickelungsgeschichte der Thiere. Beobachtung und Reflexion*. Königsberg,: Bei den Gebrüdern Borntträger,. doi: 10.5962/bhl.title.6303.

Bailey, T. L. (2011) ‘DREME: Motif discovery in transcription factor ChIP-seq data’,

Bioinformatics. doi: 10.1093/bioinformatics/btr261.

Bardet, A. F. *et al.* (2013) 'Identification of transcription factor binding sites from ChIP-seq data at high resolution', *Bioinformatics*. doi: 10.1093/bioinformatics/btt470.

Barski, A. *et al.* (2007) 'High-Resolution Profiling of Histone Methylations in the Human Genome', *Cell*. doi: 10.1016/j.cell.2007.05.009.

Bassham, S. and Postlethwait, J. (2000) 'Brachyury (T) expression in embryos of a larvacean urochordate, Oikopleura dioica, and the ancestral role of T', *Developmental Biology*, 220(2), pp. 322–332. doi: 10.1006/dbio.2000.9647.

Baxter, A. L. (1977) 'E. B. Wilson's "Destruction" of the Germ-Layer Theory', *Isis*, 68(3), pp. 363–374. doi: 10.1086/351813.

Ben-Tabou de-Leon, S. *et al.* (2013) 'Gene regulatory control in the sea urchin aboral ectoderm: Spatial initiation, signaling inputs, and cell fate lockdown', *Developmental Biology*, 374(1), pp. 245–254. doi: 10.1016/j.ydbio.2012.11.013.

Bernardo, M. D. I. *et al.* (2000) 'Homeobox genes and sea urchin development', 643, pp. 637–643.

Berns, N. *et al.* (2008) 'Expression, function and regulation of Brachyenteron in the short germband insect Tribolium castaneum', *Development Genes and Evolution*, 218(3–4), pp. 169–179. doi: 10.1007/s00427-008-0210-7.

Betts, M. J. and Russell, R. B. (2007) 'Amino-Acid Properties and Consequences of Substitutions', *Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data: Second Edition*, 4, pp. 311–342. doi: 10.1002/9780470059180.ch13.

Bhagwat, M. and Aravind, L. (2007) 'PSI-BLAST tutorial', *Methods in Molecular Biology*. doi: 10.1385/1-59745-514-8:177.

Bielen, H. *et al.* (2007) 'Divergent functions of two ancient Hydra Brachyury paralogues suggest specific roles for their C-terminal domains in tissue fate induction', *Development*. doi: 10.1242/dev.010173.

Bland, J. M. and Altman, D. G. (1995) 'Multiple significance tests: The Bonferroni

method', *BMJ*. doi: 10.1136/bmj.310.6973.170.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics*. doi: 10.1093/bioinformatics/btu170.

Booth, D. S. and King, N. (2016) 'Gene Regulation in transition', *Nature*, 543, pp. 482–3. doi: 10.1038/nature18447.

Breitling, J. and Aebi, M. (2013) 'N-linked protein glycosylation in the endoplasmic reticulum', *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a013359.

Brown T. A. (2002) 'Transcriptomes and Proteomes', in Brown T. A. (ed.) *Genomes*. Edition 2. Oxford: Wiley-Liss. Available at: https://www.ncbi.nlm.nih.gov/books/NBK21121/.

Bryne, J. C. *et al.* (2008) 'JASPAR, the open access database of transcription factor-binding profiles: New content and tools in the 2008 update', *Nucleic Acids Research*. doi: 10.1093/nar/gkm955.

Buenrostro, J. D. *et al.* (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nat Meth*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 10(12), pp. 1213–1218. Available at: http://dx.doi.org/10.1038/nmeth.2688.

Buenrostro, J. D. *et al.* (2015) 'ATAC-seq: A method for assaying chromatin accessibility genome-wide', in *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., pp. 21.29.1-21.29.9. doi: 10.1002/0471142727.mb2129s109.

Cheers, M. S. and Ettensohn, C. A. (2004) 'Rapid microinjection of fertilized eggs.', *Methods in cell biology*. United States, 74, pp. 287–310.

Chen, J.-Y. *et al.* (2004) 'Small Bilaterian Fossils from 40 to 55 Million Years Before the Cambrian', *Science*, 305(5681), pp. 218 LP – 222. doi: 10.1126/science.1099213.

Chen, J. H., Luo, Y. J. and Su, Y. H. (2011) 'The dynamic gene expression patterns of transcription factors constituting the sea urchin aboral ectoderm gene

regulatory network', *Developmental Dynamics*. doi: 10.1002/dvdy.22514.

Chen, Y. L. *et al.* (2009) 'Smad6 inhibits the transcriptional activity of Tbx6 by mediating its degradation', *Journal of Biological Chemistry*, 284(35), pp. 23481–23490. doi: 10.1074/jbc.M109.007864.

Coffman, J. A. *et al.* (1996) 'SpRunt-1, a new member of the runt domain family of transcription factors, is a positive regulator of the aboral ectoderm-specific CyIIIA gene in sea urchin embryos', *Developmental Biology*. doi: 10.1006/dbio.1996.0050.

Cole, A. G. and Arnone, M. I. (2009) 'Fluorescent in situ hybridization reveals multiple expression domains for SpBrn1/2/4 and identifies a unique ectodermal cell type that co-expresses the ParaHox gene SpLox', *Gene Expression Patterns*. doi: 10.1016/j.gep.2009.02.005.

Coley, C. *et al.* (2000) 'Effect of multiple serine/alanine mutations in the transmembrane spanning region V of the D2 dopamine receptor on ligand binding', *Journal of Neurochemistry*. doi: 10.1046/j.1471-4159.2000.0740358.x.

Collas, P. (2010) 'The current state of chromatin immunoprecipitation', *Molecular Biotechnology*. doi: 10.1007/s12033-009-9239-8.

Conlon, F. L. *et al.* (1996) 'Inhibition of Xbra transcription activation causes defects in mesodermal patterning and reveals autoregulation of Xbra in dorsal mesoderm.', *Development (Cambridge, England)*, 122(8), pp. 2427–2435. Available at: http://dev.biologists.org/content/122/8/2427.abstract.

Conlon, F. L. *et al.* (2001) 'Determinants of T box protein specificity.', *Development*, 128(19), pp. 3749–3758. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=11585801&retmode=ref&cmd=prlinks%5Cnpapers2://publication/uuid/F3E77488-7BF5-4581-A893-B34C4871CBAD.

Conticello, S. G. *et al.* (2005) 'Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msi026.

Corbo, J. C. *et al.* (1997) 'Dorsoventral patterning of the vertebrate neural tube is

conserved in a protochordate', *Development*.

Croce, J. C. and Mcclay, D. R. (2010) 'Dynamics of Delta / Notch signaling on endomesoderm segregation in the sea urchin embryo', 91, pp. 83–91. doi: 10.1242/dev.044149.

Croce, J., Lhomond, G. and Gache, C. (2001) 'Expression pattern of Brachyury in the embryo of the sea urchin Paracentrotus lividus', *Development Genes and Evolution*, 211(12), pp. 617–619. doi: 10.1007/s00427-001-0200-5.

Crum, T. L. and Okkema, P. G. (2007) 'SUMOylation-dependent function of a T-box transcriptional repressor in Caenorhabditis elegans', *Biochemical Society Transactions*, 35(6), pp. 1424–1426. doi: 10.1042/BST0351424.

Cui, M. *et al.* (2014) 'Specific functions of the Wnt signaling system in gene regulatory networks throughout the early sea urchin embryo', *Proceedings of the National Academy of Sciences*, 111(47), pp. E5029–E5038. doi: 10.1073/pnas.1419141111.

Das, P., Maduzia, L. L. and Padgett, R. W. (1999) 'Genetic approaches to TGFβ signaling pathways', *Cytokine & Growth Factor Reviews*, 10(3–4), pp. 179–186. doi: 10.1016/S1359-6101(99)00014-3.

Davidson, E. (2006) *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution*. 1st editio, *San Diego (California): Academic Press/Elsevier*. 1st editio. San Diego: Academic Press/Elsevier.

Davidson, E. H. *et al.* (2002) 'A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo', *Developmental Biology*, 246(1), pp. 162–190. doi: 10.1006/dbio.2002.0635.

Davidson, E. H. (2010) 'Emerging properties of animal gene regulatory networks', *Nature*, 468(7326), pp. 911–920. doi: 10.1038/nature09645.

Davidson, E. H. and Levine, M. (2005) 'Gene regulatory networks', 102(14), p. 4935. doi: 10.1073/pnas.0502024102.

de-Leon, S. B.-T. and Davidson, E. H. (2010) 'Information processing at the foxa node of the sea urchin endomesoderm specification network', *Proceedings of the*

National Academy of Sciences. National Academy of Sciences, 107(22), pp. 10103–10108. doi: 10.1073/pnas.1004824107.

Deng, L. *et al.* (2000) 'Activation of the IκB Kinase Complex by TRAF6 Requires a Dimeric Ubiquitin-Conjugating Enzyme Complex and a Unique Polyubiquitin Chain', *Cell*, 103(2), pp. 351–361. doi: 10.1016/S0092-8674(00)00126-4.

Dohrmann, M. and Worheide, G. (2013) 'Novel Scenarios of Early Animal Evolution--Is It Time to Rewrite Textbooks?', *Integrative and Comparative Biology*, 53(3), pp. 503–511. doi: 10.1093/icb/ict008.

Duboc, V. *et al.* (2004) 'Nodal and BMP2/4 signaling organizes the oral-aboral axis of the sea urchin embryo', *Developmental Cell*, 6(3), pp. 397–410. doi: 10.1016/S1534-5807(04)00056-5.

Duboc, V. *et al.* (2005) 'Left-Right Asymmetry in the Sea Urchin Embryo Is Regulated by Nodal Signaling on the Right Side', 9, pp. 147–158. doi: 10.1016/j.devcel.2005.05.008.

Dylus, D. V. *et al.* (2016) 'Large-scale gene expression study in the ophiuroid Amphiura filiformis provides insights into evolution of gene regulatory networks', *EvoDevo*. BioMed Central, 7(1). doi: 10.1186/s13227-015-0039-x.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*. doi: 10.1093/nar/gkh340.

Eitel, M. *et al.* (2011) 'New insights into placozoan sexual reproduction and development', *PloS one*. 2011/05/19. Public Library of Science, 6(5), pp. e19639–e19639. doi: 10.1371/journal.pone.0019639.

Erkenbrack, E. M., Davidson, E. H. and Peter, I. S. (2018) 'Conserved regulatory state expression controlled by divergent developmental gene regulatory networks in echinoids', *Development (Cambridge)*, 145(24). doi: 10.1242/dev.167288.

Erwin, D. H. and Davidson, E. H. (2009) 'The evolution of hierarchical gene regulatory networks.', *Nature reviews. Genetics*, 10(2), pp. 141–148. doi: 10.1038/nrg2499.

Ettensohn, C. A. (2003) 'Alx1, a member of the Cart1/Alx3/Alx4 subfamily of

Paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo', *Development*, 130(13), pp. 2917–2928. doi: 10.1242/dev.00511.

Evans, A. L. *et al.* (2012) 'Genomic Targets of Brachyury (T) in Differentiating Mouse Embryonic Stem Cells', 7(3). doi: 10.1371/journal.pone.0033346.

Evren, S. *et al.* (2014) 'EphA4-dependent Brachyury expression is required for dorsal mesoderm involution in the Xenopus gastrula', *Development*, 141(19), pp. 3649–3661. doi: 10.1242/dev.111880.

Faial, T. *et al.* (2015) 'Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells', *Development (Cambridge)*, 142(12), pp. 2121–2135. doi: 10.1242/dev.117838.

Feher, J. (2012) 'Protein Structure', in *Quantitative Human Physiology*. Elsevier, pp. 100–109. doi: 10.1016/B978-0-12-382163-8.00012-8.

Force, A. *et al.* (1999) 'Preservation of duplicate genes by complementary, degenerative mutations.', *Genetics*. United States, 151(4), pp. 1531–1545.

Fulton, M. D., Brown, T. and George Zheng, Y. (2019) 'The biological axis of protein arginine methylation and asymmetric dimethylarginine', *International Journal of Molecular Sciences*. doi: 10.3390/ijms20133322.

Gareau, J. R. and Lima, C. D. (2010) 'The SUMO pathway: Emerging mechanisms that shape specificity, conjugation and recognition', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 11(12), pp. 861–871. doi: 10.1038/nrm3011.

Gaszner, M. and Felsenfeld, G. (2006) 'Insulators: Exploiting transcriptional and epigenetic mechanisms', *Nature Reviews Genetics*. doi: 10.1038/nrg1925.

Gavel, Y. and Heijne, G. Von (1990) 'Sequence differences between glycosylated and non-glycosylated asn-x-thr/ser acceptor sites: Implications for protein engineering', *Protein Engineering, Design and Selection*. doi: 10.1093/protein/3.5.433.

Genikhovich, G. and Technau, U. (2017) 'On the evolution of bilaterality',

Development (Cambridge), 144(19), pp. 3392–3404. doi: 10.1242/dev.141507.

Gentsch, G. E. *et al.* (2013) 'InVivo T-Box Transcription Factor Profiling Reveals Joint Regulation of Embryonic Neuromesodermal Bipotency', *Cell Reports*. The Authors, 4(6), pp. 1185–1196. doi: 10.1016/j.celrep.2013.08.012.

Gilbert, S. F. (2003) 'The morphogenesis of evolutionary developmental biology', *International Journal of Developmental Biology*, 47(7–8), pp. 467–477. doi: 10.1387/ijdb.14756322.

Gilbert, S. F. and Barresi, M. J. F. (2017) 'DEVELOPMENTAL BIOLOGY, 11TH EDITION 2016', *American Journal of Medical Genetics Part A*, 173(5), pp. 1430–1430. doi: 10.1002/ajmg.a.38166.

Gill, G. (2004) 'SUMO and ubiquitin in the nucleus: Different functions, similar mechanisms?', *Genes and Development*. doi: 10.1101/gad.1214604.

Gill, G. (2005) 'Something about SUMO inhibits transcription', *Current Opinion in Genetics and Development*. doi: 10.1016/j.gde.2005.07.004.

Ginestet, C. (2011) 'ggplot2: Elegant Graphics for Data Analysis', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. doi: 10.1111/j.1467-985x.2010.00676_9.x.

Goel, M. and Mushegian, A. (2006) 'Intermediary metabolism in sea urchin: The first inferences from the genome sequence', *Developmental Biology*, 300(1), pp. 282–292. doi: 10.1016/j.ydbio.2006.08.030.

Green, S. and Batterman, R. (2017) 'Biology meets physics: Reductionism and multi-scale modeling of morphogenesis', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. Pergamon, 61, pp. 20–34. doi: 10.1016/J.SHPSC.2016.12.003.

Gromiha, M. M., Nagarajan, R. and Selvaraj, S. (2019) 'Protein Structural Bioinformatics: An Overview', *Encyclopedia of Bioinformatics and Computational Biology*, pp. 445–459. doi: 10.1016/b978-0-12-809633-8.20278-1.

Gross, J. M. and McClay, D. R. (2001) 'The role of Brachyury (T) during gastrulation movements in the sea urchin Lytechinus variegatus', *Dev Biol*, 239(1),

pp. 132–147. doi: 10.1006/dbio.2001.0426.

Harada, Y. *et al.* (1996) ‘Spatial expression of a forkhead homologue in the sea urchin embryo’, *Mechanisms of Development*, 60(2), pp. 163–173. doi: 10.1016/S0925-4773(96)00608-9.

Harada, Y., Yasuo, H. and Satoh, N. (1995) ‘A sea urchin homologue of the chordate Brachyury (T) gene is expressed in the secondary mesenchyme founder cells.’, *Development (Cambridge, England)*, 121(9), pp. 2747–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7555703.

Hardin, S. H. *et al.* (1985) ‘Structure of the Specl Gene Encoding a Major Calcium-binding Protein in the Embryonic Ectoderm of the Sea Urchin , Strongylocentrotus purpuratus’.

Hay, R. T. (2005) ‘SUMO: a history of modification’, *Molecular Cell*, 18(1), pp. 1–12. doi: 10.1016/j.molcel.2005.03.012.

Hayward, D. C. *et al.* (2015) ‘The organizer in evolution-gastrulation and organizer gene expression highlight the importance of Brachyury during development of the coral, Acropora millepora’, *Developmental Biology*. Elsevier, 399(2), pp. 337–347. doi: 10.1016/j.ydbio.2015.01.006.

Heinz, S. *et al.* (2010) ‘Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.’, *Molecular cell*. doi: 10.1016/j.molcel.2010.05.004.

Hejnol, A. and Martín-Durán, J. M. (2015) ‘Getting to the bottom of anal evolution’, *Zoologischer Anzeiger*. Elsevier GmbH., 256, pp. 61–74. doi: 10.1016/j.jcz.2015.02.006.

Herrmann, B. G. *et al.* (1990) ‘Cloning of the T gene required in mesoderm formation in the mouse’, *Nature*, 343(6259), pp. 617–622. doi: 10.1038/343617a0.

Hibino, T. *et al.* (2004) ‘Molecular heterotopy in the expression of Brachyury orthologs in order Clypeasteroida (irregular sea urchins) and order Echinoida (regular sea urchins)’, *Development Genes and Evolution*, 214(11), pp. 546–558. doi: 10.1007/s00427-004-0437-x.

Hofmann, R. M. and Pickart, C. M. (1999) ‘Noncanonical MMS2-Encoded Ubiquitin-Conjugating Enzyme Functions in Assembly of Novel Polyubiquitin Chains for DNA Repair’, *Cell*, 96(5), pp. 645–653. doi: 10.1016/S0092-8674(00)80575-9.

Hofmann, R. M. and Pickart, C. M. (2001) ‘In Vitro Assembly and Recognition of Lys-63 Polyubiquitin Chains’, *Journal of Biological Chemistry*. doi: 10.1074/jbc.M103378200.

Holland, P. W. *et al.* (1995) ‘Conservation of Brachyury (T) genes in amphioxus and vertebrates: developmental and evolutionary implications.’, *Development (Cambridge, England)*, 121(12), pp. 4283–4291. Available at: http://dev.biologists.org/content/121/12/4283.abstract.

Howard-Ashby, M., Materna, Stefan C., *et al.* (2006) ‘Gene families encoding transcription factors expressed in early development of Strongylocentrotus purpuratus’, *Developmental Biology*, 300(1), pp. 90–107. doi: 10.1016/j.ydbio.2006.08.033.

Howard-Ashby, M., Materna, Stefan C, *et al.* (2006) ‘Identification and characterization of homeobox transcription factor genes in Strongylocentrotus purpuratus, and their expression in embryonic development’, *Developmental Biology*, 300(1), pp. 74–89. doi: 10.1016/j.ydbio.2006.08.039.

Huber, P. *et al.* (2013) ‘Function of the C. elegans T-box factor TBX-2 depends on SUMOylation’, *Cellular and Molecular Life Sciences*, 70(21), pp. 4157–4168. doi: 10.1007/s00018-013-1336-y.

Hübner, M. R., Eckersley-Maslin, M. A. and Spector, D. L. (2013) ‘Chromatin organization and transcriptional regulation’, *Current Opinion in Genetics and Development*. doi: 10.1016/j.gde.2012.11.006.

Ignatiadis, N. *et al.* (2016) ‘Data-driven hypothesis weighting increases detection power in genome-scale multiple testing’, *Nature Methods*. doi: 10.1038/nmeth.3885.

Ignatova, Z. and Gierasch, L. M. (2009) ‘A Method for Direct Measurement of Protein Stability In Vivo’, in, pp. 165–178. doi: 10.1007/978-1-59745-367-7_7.

Ikeda, T. and Satou, Y. (2017) ‘ Differential temporal control of Foxa.a and Zic-r.b specifies brain versus notochord fate in the ascidian embryo ’, *Development*, 144(1), pp. 38–43. doi: 10.1242/dev.142174.

Ilari, A. and Savino, C. (2008) ‘Protein Structure Determination by X-Ray Crystallography’, in, pp. 63–87. doi: 10.1007/978-1-60327-159-2_3.

Illies, M. R. *et al.* (2002) ‘Cloning and developmental expression of a novel, secreted frizzled-related protein from the sea urchin, Strongylocentrotus purpuratus’, *Mechanisms of Development*, 113(1), pp. 61–64. doi: 10.1016/S0925-4773(01)00657-8.

Israel, J. W. *et al.* (2016) ‘Comparative Developmental Transcriptomics Reveals Rewiring of a Highly Conserved Gene Regulatory Network during a Major Life History Switch in the Sea Urchin Genus Heliocidaris’, *PLoS Biology*, 14(3), pp. 1–28. doi: 10.1371/journal.pbio.1002391.

Jager, M. and Manuel, M. (2016) ‘Ctenophores: an evolutionary-developmental perspective’, *Current Opinion in Genetics and Development*, 39, pp. 85–92. doi: 10.1016/j.gde.2016.05.020.

Jang, E. J. *et al.* (2013) ‘Lysine 313 of T-box Is Crucial for Modulation of Protein Stability, DNA Binding, and Threonine Phosphorylation of T-bet’, *The Journal of Immunology*, 190(11), pp. 5764–5770. doi: 10.4049/jimmunol.1203403.

José-Edwards, D. S. *et al.* (2015) ‘Brachyury, Foxa2 and the cis-Regulatory Origins of the Notochord’, *PLoS Genetics*, 11(12). doi: 10.1371/journal.pgen.1005730.

Joshi, R., Sun, L. and Mann, R. (2010) ‘Dissecting the functional specificities of two Hox proteins’, *Genes and Development*. doi: 10.1101/gad.1936910.

Kalhan, S. C. and Hanson, R. W. (2012) ‘Resurgence of serine: An often neglected but indispensable amino acid’, *Journal of Biological Chemistry*, 287(24), pp. 19786–19791. doi: 10.1074/jbc.R112.357194.

Källberg, M. *et al.* (2012) ‘Template-based protein structure modeling using the RaptorX web server’, *Nature Protocols*. doi: 10.1038/nprot.2012.085.

Katikala, L. *et al.* (2013) ‘Functional Brachyury Binding Sites Establish a Temporal

Read-out of Gene Expression in the Ciona Notochord', *PLoS Biology*, 11(10). doi: 10.1371/journal.pbio.1001697.

Khmelinskii, A. *et al.* (2012) 'Tandem fluorescent protein timers for in vivo analysis of protein dynamics.', *Nature biotechnology*. Nature Publishing Group, 30(7), pp. 708–14. doi: 10.1038/nbt.2281.

Khor, J. M. and Ettensohn, C. A. (2017) 'Functional divergence of paralogous transcription factors supported the evolution of biomineralization in echinoderms', *eLife*, 6, pp. 1–21. doi: 10.7554/eLife.32728.

Khor, J. M., Guerrero-Santoro, J. and Ettensohn, C. A. (2019) 'Genome-wide identification of binding sites and gene targets of Alx1, a pivotal regulator of echinoderm skeletogenesis', *Development*, (July), p. dev.180653. doi: 10.1242/dev.180653.

King, N. and Rokas, A. (2017) 'Embracing Uncertainty in Reconstructing Early Animal Evolution', *Current Biology*. Elsevier Ltd, 27(19), pp. R1081–R1088. doi: 10.1016/j.cub.2017.08.054.

Kispert, A. *et al.* (1994) 'Homologs of the mouse Brachyury gene are involved in the specification of posterior terminal structures in Drosophila, Tribolium, and Locusta', *Genes and Development*, 8(18), pp. 2137–2150. doi: 10.1101/gad.8.18.2137.

Kispert, A. and Herrmann, B. G. (1993) 'The Brachyury gene encodes a novel DNA binding protein.', *The EMBO journal*, 12(8), pp. 3211–3220. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8344258%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/8344258.

Kispert, A., Koschorz, B. and Herrmann, B. G. (1995) 'The T protein encoded by Brachyury is a tissue-specific transcription factor.', *The EMBO journal*, 14(19), pp. 4763–72. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7588606%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC394574.

Kolovos, P. *et al.* (2012) 'Enhancers and silencers: An integrated and simple model for their function', *Epigenetics and Chromatin*. doi: 10.1186/1756-8935-5-1.

Kotake, Y. *et al.* (2005) 'Role of serine 10 phosphorylation in p27 stabilization revealed by analysis of p27 knock-in mice harboring a serine 10 mutation', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M406117200.

Krishna, R. G. and Wold, F. (2006) 'Post-Translational Modification of Proteins', in *Advances in Enzymology and Related Areas of Molecular Biology*. doi: 10.1002/9780470123133.ch3.

Kudtarkar, P. and Cameron, R. A. (2017) 'Echinobase: an expanding resource for echinoderm genomic information', *Database : the journal of biological databases and curation*. doi: 10.1093/database/bax074.

Kukurba, K. R. and Montgomery, S. B. (2015) 'RNA Sequencing and Analysis Kimberly', *Cold Spring Harbor protocols*, 2015(11), pp. 951–69. doi: 10.1101/pdb.top084970.RNA.

Kumar, Sudhir *et al.* (2018) 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msy096.

Kumar, Shiv *et al.* (2018) 'Xbra and Smad-1 cooperate to activate the transcription of neural repressor ventx1.1 in Xenopus embryos', *Scientific Reports*. Springer US, 8(1), pp. 1–11. doi: 10.1038/s41598-018-29740-9.

Kusch, T. and Reuter, R. (1999) 'Functions for Drosophila brachyenteron and forkhead in mesoderm specification and cell signalling.', *Development (Cambridge, England)*, 126(18), pp. 3991–4003. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10457009.

Langmead and Steven L Salzberg (2013) 'Bowtie2', *Nature methods*. doi: 10.1038/nmeth.1923.Fast.

Lapébie, P. *et al.* (2014) 'Differential Responses to Wnt and PCP Disruption Predict Expression and Developmental Function of Conserved and Novel Genes in a Cnidarian', *PLoS Genetics*, 10(9). doi: 10.1371/journal.pgen.1004590.

Lapraz, F., Besnardeau, L. and Lepage, T. (2009) 'Patterning of the dorsal-ventral axis in echinoderms: Insights into the evolution of the BMP-chordin signaling

network', *PLoS Biology*, 7(11). doi: 10.1371/journal.pbio.1000248.

Lartillot, N. *et al.* (2002) 'Expression pattern of Brachyury in the mollusc Patella vulgata suggests a conserved role in the establishment of the AP axis in Bilateria.', *Development (Cambridge, England)*, 129(6), pp. 1411–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11880350.

Latchman, D. S. (1997) 'Transcription factors: An overview', *The International Journal of Biochemistry & Cell Biology*, 29(12), pp. 1305–1312. doi: 10.1016/S1357-2725(97)00085-X.

Lei, H. *et al.* (2010) 'A widespread distribution of genomic cemyod binding sites revealed and cross validated by ChIP-CHip and ChIP-seq techniques', *PLoS ONE*. doi: 10.1371/journal.pone.0015898.

Leininger, S. *et al.* (2014) 'Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans', *Nature Communications*, 5(May), pp. 1–15. doi: 10.1038/ncomms4905.

Lengyel, J. A. and Iwaki, D. D. (2002) 'It Takes Guts : The Drosophila Hindgut as a Model System for Organogenesis', 19, pp. 1–19. doi: 10.1006/dbio.2002.0577.

Lepage, T. and Gache, C. (2004) 'Expression of exogenous mRNAs to study gene function in the sea urchin embryo.', *Methods in cell biology*. United States, 74, pp. 677–697.

Li A, Gao X, Ren J, Jin C, X. Y. (2009) 'BDM-PUB: computational prediction of protein ubiquitination sites with a Bayesian discriminant method'.

Li, E. and Davidson, E. H. (2009) 'Building developmental gene regulatory networks', *Birth Defects Research Part C: Embryo*, 87(2), pp. 1–13. doi: 10.1002/bdrc.20152.Building.

Li, E., Materna, S. C. and Davidson, E. H. (2012) 'Direct and indirect control of oral ectoderm regulatory gene expression by Nodal signaling in the sea urchin embryo', *Developmental Biology*, 369(2), pp. 377–385. doi: 10.1016/j.ydbio.2012.06.022.

Li, H. *et al.* (2009) 'The Sequence Alignment / Map (SAM) Format and SAMtools

1000 Genome Project Data Processing Subgroup', *Bioinformatics (Oxford, England)*. doi: 10.1093/bioinformatics/btp352.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*. doi: 10.1093/bioinformatics/btp698.

Lindeman, L. C. *et al.* (2009) 'Fish'n ChIPs: Chromatin immunoprecipitation in the Zebrafish embryo', *Methods in Molecular Biology*. doi: 10.1007/978-1-60327-414-2_5.

Livi, C. B. and Davidson, E. H. (2007) 'Regulation of spblimp1/krox1a, an alternatively transcribed isoform expressed in midgut and hindgut of the sea urchin gastrula', *Gene Expression Patterns*, 7(1–2), pp. 1–7. doi: 10.1016/j.modgep.2006.04.009.

Longabaugh, W. J. R. (2012) *Gene Regulatory Networks*. Edited by B. Deplancke and N. Gheldof. Totowa, NJ: Humana Press (Methods in Molecular Biology). doi: 10.1007/978-1-61779-292-2.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.', *Genome biology*. doi: 10.1186/s13059-014-0550-8.

Lowe, E. K., Cuomo, C. and Arnone, M. I. (2016) 'Dynamics of Mathematical Models in Biology', *Dynamics of Mathematical Models in Biology*, (December). doi: 10.1007/978-3-319-45723-9.

Lowe, E. K., Cuomo, C. and Arnone, M. I. (2017) 'Omics approaches to study gene regulatory networks for development in echinoderms', *Briefings in Functional Genomics*, 16(5), pp. 299–308. doi: 10.1093/bfgp/elx012.

Luo, Y. J. and Su, Y. H. (2012) 'Opposing Nodal and BMP Signals Regulate Left-Right Asymmetry in the Sea Urchin Larva', *PLoS Biology*, 10(10). doi: 10.1371/journal.pbio.1001402.

Ma, J. *et al.* (2012) 'A conditional neural fields model for protein threading', *Bioinformatics*. doi: 10.1093/bioinformatics/bts213.

Ma, J. *et al.* (2013) 'Protein threading using context-specific alignment potential', in

Bioinformatics. doi: 10.1093/bioinformatics/btt210.

Maeda, R. K. and Karch, F. (2011) ‘Gene expression in time and space: Additive vs hierarchical organization of cis-regulatory regions’, *Current Opinion in Genetics and Development*. doi: 10.1016/j.gde.2011.01.021.

Magri, M. S. *et al.* (2021) ‘ATAC-Seq for Assaying Chromatin Accessibility Protocol Using Echinoderm Embryos’, in Carroll, D. J. and Stricker, S. A. (eds) *Developmental Biology of the Sea Urchin and Other Marine Invertebrates: Methods and Protocols*. New York, NY: Springer US, pp. 253–265. doi: 10.1007/978-1-0716-0974-3_16.

Mahony, S., Auron, P. E. and Benos, P. V. (2007) ‘DNA familial binding profiles made easy: Comparison of various motif alignment and clustering strategies’, *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.0030061.

Mahony, S. and Benos, P. V. (2007) ‘STAMP: A web tool for exploring DNA-binding motif similarities’, *Nucleic Acids Research*. doi: 10.1093/nar/gkm272.

Marcellini, S. *et al.* (2003) ‘Evolution of Brachyury proteins: Identification of a novel regulatory domain conserved within Bilateria’, *Developmental Biology*, 260(2), pp. 352–361. doi: 10.1016/S0012-1606(03)00244-6.

Marcellini, S. (2006) ‘When Brachyury meets Smad1: The evolution of bilateral symmetry during gastrulation’, *BioEssays*, 28(4), pp. 413–420. doi: 10.1002/bies.20387.

Marsh, A. G., Leong, P. K. K. and Manahan, D. T. (1999) ‘Energy metabolism during embryonic development and larval growth of an Antarctic sea urchin’, *Journal of Experimental Biology*, 202(15), pp. 2041–2050.

Martín-Durán, J. M. *et al.* (2016) ‘The developmental basis for the recurrent evolution of deuterostomy and protostomy’, *Nature Ecology & Evolution*. Nature Publishing Group, 1(1), p. 0005. doi: 10.1038/s41559-016-0005.

Martindale, M. Q., Pang, K. and Finnerty, J. R. (2004) ‘Investigating the origins of triplosblasty: “Mesodermal” gene expression in a diploblastic animal, the sea anemone Nematostella vectensis (phylum, Cnidaria; class, Anthozoa)’,

Development, 131(10), pp. 2463–2474. doi: 10.1242/dev.01119.

Martinelli, C. and Spring, J. (2003) ‘Distinct expression patterns of the two T-box homologues Brachyury and Tbx2/3 in the placozoan Trichoplax adhaerens’, *Development Genes and Evolution*, 213(10), pp. 492–499. doi: 10.1007/s00427-003-0353-5.

Masabni, J. G. and Zandstra, B. H. (1999) ‘ A serine-to-threonine mutation in linuron-resistant Portulaca oleracea ’, *Weed Science*. doi: 10.1017/s0043174500091979.

Massagué, J., Seoane, J. and Wotton, D. (2005) ‘Smad transcription factors’, *Genes and Development*, 19(23), pp. 2783–2810. doi: 10.1101/gad.1350705.

Materna, S. C. *et al.* (2006) ‘The C 2 H 2 zinc finger genes of Strongylocentrotus purpuratus and their expression in embryonic development’, 300, pp. 108–120. doi: 10.1016/j.ydbio.2006.08.032.

Materna, S. C. (2017) ‘Using Morpholinos to Probe Gene Networks in Sea Urchin’, in, pp. 87–104. doi: 10.1007/978-1-4939-6817-6_8.

Materna, S. C. and Davidson, E. H. (2012) ‘A comprehensive analysis of Delta signaling in pre-gastrular sea urchin embryos’, *Developmental Biology*, 364(1), pp. 77–87. doi: 10.1016/j.ydbio.2012.01.017.

Materna, S. C., Swartz, S. Z. and Smith, J. (2013) ‘Notch and Nodal control forkhead factor expression in the specification of multipotent progenitors in sea urchin’, 1806, pp. 1796–1806. doi: 10.1242/dev.091157.

Matys, V. *et al.* (2003) ‘TRANSFAC®: Transcriptional regulation, from patterns to profiles’, *Nucleic Acids Research*. doi: 10.1093/nar/gkg108.

McClay, D. R. *et al.* (2000) ‘A micromere induction signal is activated by ??-catenin and acts through Notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo’, *Development*.

McClay, D. R. (2011) ‘Evolutionary crossroads in developmental biology: sea urchins.’, *Development (Cambridge, England)*, 138(13), pp. 2639–48. doi: 10.1242/dev.048967.

McClay, D. R., Miranda, E. and Feinberg, S. L. (2018) ‘Neurogenesis in the sea urchin embryo is initiated uniquely in three domains’, *Development (Cambridge, England)*. doi: 10.1242/dev.167742.

Mcclay, R. and Ettensohn, C. (1988) ‘Cell lineage conversion in the Sea Urchin Embryo’, *Developmental Biology*, 409, pp. 396–409.

Mcintyre, D. C. *et al.* (2013) ‘Short-range Wnt5 signaling initiates specification of sea urchin posterior ectoderm’, *Development (Cambridge)*, 140(24), pp. 4881–4889. doi: 10.1242/dev.095844.

Mende, M., Christophorou, N. A. D. and Streit, A. (2008) ‘Specific and effective gene knock-down in early chick embryos using morpholinos but not pRFPRNAi vectors’, *Mechanisms of Development*, 125(11–12), pp. 947–962. doi: 10.1016/j.mod.2008.08.005.

Mendoza, L., Taylor, J. W. and Ajello, L. (2002) ‘The Class Mesomycetozoea: A Heterogeneous Group of Microorganisms at the Animal-Fungal Boundary’, *Annual Review of Microbiology*, 56(1), pp. 315–344. doi: 10.1146/annurev.micro.56.012302.160950.

Messenger, N. J. *et al.* (2005) ‘Functional specificity of the xenopus T-domain protein brachyury is conferred by its ability to interact with smad1’, *Developmental Cell*, 8(4), pp. 599–610. doi: 10.1016/j.devcel.2005.03.001.

Mi, H. *et al.* (2016) ‘Protocol Update for large-scale genome and gene function analysis with the PANTHER’, *Nature Protocols*. Springer US. doi: 10.1038/s41596-019-0128-8.

Mi, H. *et al.* (2019) ‘Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0)’, *Nature Protocols*. doi: 10.1038/s41596-019-0128-8.

Milton, A. C. and Okkema, P. G. (2015) ‘Caenorhabditis elegans TBX-2 directly regulates its own expression in a negative autoregulatory loop’, *G3: Genes, Genomes, Genetics*, 5(6), pp. 1177–1186. doi: 10.1534/g3.115.018101.

Minokawa, T. *et al.* (2004) ‘Expression patterns of four different regulatory genes

that function during sea urchin development', 4, pp. 449–456. doi: 10.1016/j.modgep.2004.01.009.

Momose, T., Derelle, R. and Houliston, E. (2008) 'A maternally localised Wnt ligand required for axial patterning in the cnidarian Clytia hemisphaerica', *Development*, 135(12), pp. 2105–2113. doi: 10.1242/dev.021543.

Müller, C. W. *et al.* (1997) 'Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor.', *Nature*. England, 389(6653), pp. 884–888. doi: 10.1038/39929.

Müller, W. E. G. (2001) 'Review: How was metazoan threshold crossed? The hypothetical Urmetazoa', *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*. doi: 10.1016/S1095-6433(00)00360-3.

Nakatani, Y. *et al.* (1996) 'Basic fibroblast growth factor induces notochord formation and the expression of As-T, a Brachyury homolog, during ascidian embryogenesis.', *Development (Cambridge, England)*, 122(7), pp. 2023–31. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8681783.

Oliveri, P. *et al.* (2006) 'Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo', *Development*, 133(21), pp. 4173–4181. doi: 10.1242/dev.02577.

Oliveri, P., Carrick, D. M. and Davidson, E. H. (2002) 'A Regulatory Gene Network That Directs Micromere Specification in the Sea Urchin Embryo', *Developmental Biology*, 246(1), pp. 209–228. doi: 10.1006/dbio.2002.0627.

Onai, T. *et al.* (2009) 'Retinoic acid and Wnt/β-catenin have complementary roles in anterior/posterior patterning embryos of the basal chordate amphioxus', *Developmental Biology*. Elsevier Inc., 332(2), pp. 223–233. doi: 10.1016/j.ydbio.2009.05.571.

Oppenheimer, J. M. (1990) 'Science and Nationality: The Case of Karl Ernst von Baer (1792-1876)', *Proceedings of the American Philosophical Society*. American Philosophical Society, 134(2), pp. 75–82. Available at: http://www.jstor.org/stable/986838.

Orlando, V. (2000) ‘Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation’, *Trends in Biochemical Sciences*. doi: 10.1016/S0968-0004(99)01535-2.

Otim, O. *et al.* (2004) ‘SpHnf6 , a transcription factor that executes multiple functions in sea urchin embryogenesis’, 273, pp. 226–243. doi: 10.1016/j.ydbio.2004.05.033.

Pan, L. *et al.* (2014) ‘Deubiquitination and stabilization of T-bet by USP10’, *Biochemical and Biophysical Research Communications*. Elsevier Inc., 449(3), pp. 289–294. doi: 10.1016/j.bbrc.2014.05.037.

Papaioannou, V. E. (2001) ‘T-box genes in development: From hydra to humans’, *International Review of Cytology*, 207, pp. 1–70. doi: 10.1016/S0074-7696(01)07002-4.

Papaioannou, V. E. (2014) ‘The T-box gene family: emerging roles in development, stem cells and cancer’, *Development*, 141(20), pp. 3819–3833. doi: 10.1242/dev.104471.

Paps, J., Holland, P. W. H. and Shimeld, S. M. (2012) ‘A genome-wide view of transcription factor gene diversity in chordate evolution: Less gene loss in amphioxus?’, *Briefings in Functional Genomics*. doi: 10.1093/bfgp/els012.

Parekh, R. B. and Rohlff, C. (1997) ‘Post-translational modification of proteins and the discovery of new medicine’, *Current Opinion in Biotechnology*. doi: 10.1016/S0958-1669(97)80126-7.

Passamore, L. A. and Barford, D. (2004) ‘Getting into position: the catalytic mechanisms of protein ubiquitylation’, *Biochemical Journal*, 379(3), pp. 513–525. doi: 10.1042/bj20040198.

Patro, R. *et al.* (2017) ‘Salmon provides fast and bias-aware quantification of transcript expression’, *Nature Methods*. Nature Publishing Group, 14(4), pp. 417–419. doi: 10.1038/nmeth.4197.

Peng, J. and Xu, J. (2011) ‘A multiple-template approach to protein threading’, *Proteins: Structure, Function and Bioinformatics*. doi: 10.1002/prot.23016.

Peres, J., Mowla, S. and Prince, S. (2015) 'The T-box transcription factor, TBX3, is a key substrate of AKT3 in melanomagenesis', *Oncotarget*. doi: 10.18632/oncotarget.2782.

Pérez-Pomares, J. M. and Muñoz-Chápuli, R. (2002) 'Epithelial-mesenchymal transitions: A mesodermal cell strategy for evolutive innovation in Metazoans', *The Anatomical Record*, 268(3), pp. 343–351. doi: 10.1002/ar.10165.

Perillo, M. *et al.* (2016) 'A pancreatic exocrine-like cell regulatory circuit operating in the upper stomach of the sea urchin Strongylocentrotus purpuratus larva', *BMC Evolutionary Biology*. BMC Evolutionary Biology, 16(1), pp. 1–15. doi: 10.1186/s12862-016-0686-0.

Perillo, M. *et al.* (2018) 'New Neuronal Subtypes With a " Pre-Pancreatic " Signature in the Sea Urchin Stongylocentrotus purpuratus', 9(November), pp. 1–12. doi: 10.3389/fendo.2018.00650.

Peter, I. and Davidson, E. (2015) *Genomic control process: development and evolution*.

Peter, I. S. (2017) 'Regulatory states in the developmental control of gene expression', *Briefings in Functional Genomics*, 16(5), pp. 281–287. doi: 10.1093/bfgp/elx009.

Peter, I. S. and Davidson, E. H. (2011a) 'A gene regulatory network controlling the embryonic specification of endoderm', *Nature*, 474(7353), pp. 635–639. doi: 10.1038/nature10100.

Peter, I. S. and Davidson, E. H. (2011b) 'Evolution of Gene Regulatory Networks Controlling Body Plan Development', *Cell*, 144(6), pp. 970–985. doi: 10.1016/j.cell.2011.02.017.

Peterson, K. J., Cameron, R. A., *et al.* (1999) 'A comparative molecular approach to mesodermal patterning in basal deuterostomes: the expression pattern of Brachyury in the enteropneust hemichordate Ptychodera flava.', *Development (Cambridge, England)*, 126(1), pp. 85–95. Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9834188&retmode=ref&cmd=prlinks%5Cnpapers3://publication/uuid/900CBED5-C39A-

45FD-B38A-6CAD9B328D35.

Peterson, K. J., Harada, Y., *et al.* (1999) ‘Expression pattern of Brachyury and Not in the sea urchin: comparative implications for the origins of mesoderm in the basal deuterostomes.’, *Developmental biology*, 207(2), pp. 419–431. doi: 10.1006/dbio.1998.9177.

Petit, V. *et al.* (2009) ‘Murine APOBEC1 Is a Powerful Mutator of Retroviral and Cellular RNA In Vitro and In Vivo’, *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2008.10.043.

Poustka, A. J. *et al.* (2007) ‘A global view of gene expression in lithium and zinc treated sea urchin embryos: New components of gene regulatory networks’, *Genome Biology*, 8(5). doi: 10.1186/gb-2007-8-5-r85.

Prajapati, R. S. *et al.* (2007) ‘Thermodynamic effects of proline introduction on protein stability’, *Proteins: Structure, Function and Genetics*. doi: 10.1002/prot.21215.

Quinlan, A. R. and Hall, I. M. (2010) ‘BEDTools: A flexible suite of utilities for comparing genomic features’, *Bioinformatics*. doi: 10.1093/bioinformatics/btq033.

Radivojac, P. *et al.* (2010) ‘Identification, analysis, and prediction of protein ubiquitination sites’, *Proteins: Structure, Function and Bioinformatics*. doi: 10.1002/prot.22555.

Rafiq, K. *et al.* (2014) ‘Genome-wide analysis of the skeletogenic gene regulatory network of sea urchins’, *Development (Cambridge)*. doi: 10.1242/dev.105585.

Ram, P. T., Mendelsohn, J. and Mills, G. B. (2012) ‘Bioinformatics and systems biology’, *Molecular Oncology*. Elsevier B.V, 6(2), pp. 147–154. doi: 10.1016/j.molonc.2012.01.008.

Range, R. C. and Wei, Z. (2017) ‘Correction: An anterior signaling center patterns and sizes the anterior neuroectoderm of the sea urchin embryo’, *Development*, 144(8), pp. 1579–1579. doi: 10.1242/dev.152140.

Ransick, A. *et al.* (2002) ‘New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array

hybridization', *Developmental Biology*, 246(1), pp. 132–147. doi: 10.1006/dbio.2002.0607.

Rast, J. P. *et al.* (2002) 'brachyury Target genes in the early sea urchin embryo isolated by differential macroarray screening.', *Developmental biology*, 246(January), pp. 191–208. doi: 10.1006/dbio.2002.0654.

Ren, J. *et al.* (2009) 'Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0', *Proteomics*. doi: 10.1002/pmic.200800646.

Rho, H. K. and McClay, D. R. (2011) 'The control of foxN2/3 expression in sea urchin embryos and its function in the skeletogenic gene regulatory network', *Development*, 138(5), pp. 937–945. doi: 10.1242/dev.058396.

Rizzo, F. *et al.* (2006) 'Identification and developmental expression of the ets gene family in the sea urchin (Strongylocentrotus purpuratus)', 300, pp. 35–48. doi: 10.1016/j.ydbio.2006.08.012.

Rizzo, F., Coffman, J. A. and Arnone, M. I. (2016) 'An Elk transcription factor is required for Runx-dependent survival signaling in the sea urchin embryo', *Developmental Biology*. Elsevier, (2002), pp. 1–14. doi: 10.1016/j.ydbio.2016.05.026.

Robertson, A. J. *et al.* (2002) 'The expression of SpRunt during sea urchin embryogenesis', *Mechanisms of Development*. doi: 10.1016/S0925-4773(02)00201-0.

Robinson, J. T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology*. doi: 10.1038/nbt.1754.

Rodriguez, M. S., Dargemont, C. and Hay, R. T. (2001) 'SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M009476200.

Röttinger, E. *et al.* (2008) 'FGF signals guide migration of mesenchymal cells, control skeletal morphogenesis and regulate gastrulation during sea urchin development', *Development*. doi: 10.1242/dev.014282.

Roy Chowdhuri, S. *et al.* (2006) 'The T-box factor TBX-2 and the SUMO

conjugating enzyme UBC-9 are required for ABa-derived pharyngeal muscle in C. elegans', *Developmental Biology*, 295(2), pp. 664–677. doi: 10.1016/j.ydbio.2006.04.001.

Russo, R. *et al.* (2014) 'The newly characterized Pl -jun is specifically expressed in skeletogenic cells of the Paracentrotus lividus sea urchin embryo', 281, pp. 3828–3843. doi: 10.1111/febs.12911.

Sampson, D. A., Wang, M. and Matunis, M. J. (2001) 'The Small Ubiquitin-like Modifier-1 (SUMO-1) Consensus Sequence Mediates Ubc9 Binding and is Essential for SUMO-1 Modification', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M100006200.

Saudemont, A. *et al.* (2010) 'Ancestral regulatory circuits governing ectoderm patterning downstream of nodal and BMP2/4 revealed by gene regulatory network analysis in an echinoderm', *PLoS Genetics*, 6(12), pp. 1–31. doi: 10.1371/journal.pgen.1001259.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis', *Nature Methods*. doi: 10.1038/nmeth.2089.

Schnell, J. D. and Hicke, L. (2003) 'Non-traditional Functions of Ubiquitin and Ubiquitin-binding Proteins', *Journal of Biological Chemistry*, 278(38), pp. 35857–35860. doi: 10.1074/jbc.R300018200.

Scholz, C. B. and Technau, U. (2003) 'The ancestral role of Brachyury: expression of NemBra1 in the basal cnidarian Nematostella vectensis (Anthozoa).', *Development genes and evolution*, 212(12), pp. 563–570. doi: 10.1007/s00427-002-0272-x.

Schulte-Merker, S. *et al.* (1992) 'The protein product of the zebrafish homologue of the mouse T gene is expressed in nuclei of the germ ring and the notochord of the early embryo.', *Development (Cambridge, England)*, 116(4), pp. 1021–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1295726.

Sebé-Pedrós, A. *et al.* (2011) 'Unexpected repertoire of metazoan transcription factors in the unicellular holozoan capsaspora owczarzaki', *Molecular Biology and Evolution*, 28(3), pp. 1241–1254. doi: 10.1093/molbev/msq309.

Sebé-Pedrós, A., Ariza-Cosano, A., *et al.* (2013) 'Early evolution of the T-box transcription factor family.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(40), pp. 16050–5. doi: 10.1073/pnas.1309748110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1309748110.

Sebé-Pedrós, A., Irimia, M., *et al.* (2013) 'Regulated aggregative multicellularity in a close unicellular relative of metazoa', *eLife*, 2013(2), pp. 1–20. doi: 10.7554/eLife.01287.

Sebé-Pedrós, A. *et al.* (2016) 'The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity', *Cell*, 165(5), pp. 1224–1237. doi: 10.1016/j.cell.2016.03.034.

Sebé-Pedrós, A., Degnan, B. M. and Ruiz-Trillo, I. (2017) 'The origin of Metazoa: a unicellular perspective', *Nature Reviews Genetics*, 18(8), pp. 498–512. doi: 10.1038/nrg.2017.21.

Sebé-Pedrós, A. and Ruiz-Trillo, I. (2017) 'Evolution and Classification of the T-Box Transcription Factor Family', in *Current Topics in Developmental Biology*, pp. 1–26. doi: 10.1016/bs.ctdb.2016.06.004.

Serfling, E., Jasin, M. and Schaffner, W. (1985) 'Enhancers and eukaryotic gene transcription', *Trends in Genetics*. doi: 10.1016/0168-9525(85)90088-5.

Servetnick, M. D. *et al.* (2017) 'Cas9-mediated excision of *Nematostella brachyury* disrupts endoderm development, pharynx formation and oral-aboral patterning', *Development*, 144(16), pp. 2951–2960. doi: 10.1242/dev.145839.

Sharma, T. and Ettensohn, C. A. (2010) 'Activation of the skeletogenic gene regulatory network in the early sea urchin embryo', *Development*, 137(7), pp. 1149–1157. doi: 10.1242/dev.048652.

Shi, Y. and Massagué, J. (2003) 'Mechanisms of TGF-β signaling from cell membrane to the nucleus', *Cell*. doi: 10.1016/S0092-8674(03)00432-X.

Shinmyo, Y. *et al.* (2006) 'brachyenteron is necessary for morphogenesis of the posterior gut but not for anteroposterior axial elongation from the posterior growth

zone in the intermediate-germband cricket Gryllus bimaculatus', *Development*, 133(22), pp. 4539–4547. doi: 10.1242/dev.02646.

Shoguchi, E., Satoh, N. and Maruyama, Y. K. (1999) 'Pattern of Brachyury gene expression in starfish embryos resembles that of hemichordate embryos but not of sea urchin embryos', *Mechanisms of Development*, 82(1–2), pp. 185–189. doi: 10.1016/S0925-4773(99)00008-8.

Showell, C., Binder, O. and Conlon, F. L. (2004) 'T-box Genes in Early Embryogenesis', *Developmental Dynamics*, 229(1), pp. 201–218. doi: 10.1002/dvdy.10480.

Siggers, T. *et al.* (2011) 'Non-DNA-binding cofactors enhance DNA-binding specificity of a transcriptional regulatory complex', *Molecular Systems Biology*. doi: 10.1038/msb.2011.89.

Sikorski, T. W. and Buratowski, S. (2009) 'The basal initiation machinery: beyond the general transcription factors', *Current Opinion in Cell Biology*. doi: 10.1016/j.ceb.2009.03.006.

Singh, S. *et al.* (2018) 'Foxi3 transcription factor activity is mediated by a C-terminal transactivation domain and regulated by the Protein Phosphatase 2A (PP2A) complex', *Scientific Reports*. doi: 10.1038/s41598-018-35390-8.

Slattery, M. *et al.* (2011) 'Cofactor binding evokes latent differences in DNA binding specificity between hox proteins', *Cell*. doi: 10.1016/j.cell.2011.10.053.

Slota, L. A., Miranda, E. M. and McClay, D. R. (2019) 'Spatial and temporal patterns of gene expression during neurogenesis in the sea urchin Lytechinus variegatus', *EvoDevo*. BioMed Central, 10(1), pp. 1–16. doi: 10.1186/s13227-019-0115-8.

Smeda, R. J. *et al.* (1993) 'A serine-to-threonine substitution in the triazine herbicide-binding protein in potato cells results in atrazine resistance without impairing productivity', *Plant Physiology*. doi: 10.1104/pp.103.3.911.

Smith, C. L. *et al.* (2014) 'Novel Cell Types, Neurosecretory Cells, and Body Plan of the Early-Diverging Metazoan Trichoplax adhaerens', *Current Biology*, 24(14), pp. 1565–1572. doi: 10.1016/j.cub.2014.05.046.

Smith, J. (1999) ‘T-box genes: What they do and how they do it’, *Trends in Genetics*, 15(4), pp. 154–158. doi: 10.1016/S0168-9525(99)01693-5.

Smith, J. *et al.* (2008) ‘A spatially dynamic cohort of regulatory genes in the endomesodermal gene network of the sea urchin embryo’, *Developmental Biology*, 313(2), pp. 863–875. doi: 10.1016/j.ydbio.2007.10.042.

Sodergren, E. *et al.* (2006) ‘The genome of the sea urchin Strongylocentrotus purpuratus’, *Science*, 314(5801), pp. 941–952. doi: 10.1126/science.1133609.The.

Solek, C. M. *et al.* (2013) ‘An ancient role for Gata-1/2/3 and Scl transcription factor homologs in the development of immunocytes’, *Developmental Biology*. Elsevier, 382(1), pp. 280–292. doi: 10.1016/j.ydbio.2013.06.019.

Solomon, M. J., Larsen, P. L. and Varshavsky, A. (1988) ‘Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene’, *Cell*. doi: 10.1016/S0092-8674(88)90469-2.

Soneson, C., Love, M. I. and Robinson, M. D. (2016) ‘Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences’, *F1000Research*. doi: 10.12688/f1000research.7563.2.

Song, Q. and Smith, A. D. (2011) ‘Identifying dispersed epigenomic domains from ChIP-seq data’, *Bioinformatics*. doi: 10.1093/bioinformatics/btr030.

Sorrells, T. R. and Johnson, A. D. (2015) ‘Making sense of transcription networks’, *Cell*. Elsevier Inc., 161(4), pp. 714–723. doi: 10.1016/j.cell.2015.04.014.

Spence, J. *et al.* (2000) ‘Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain’, *Cell*. doi: 10.1016/S0092-8674(00)00011-8.

Spring, J. *et al.* (2002) ‘Conservation of Brachyury, Mef2, and Snail in the myogenic lineage of jellyfish: A connection to the mesoderm of bilateria’, *Developmental Biology*, 244(2), pp. 372–384. doi: 10.1006/dbio.2002.0616.

Sprinkle, J. (1992) ‘Radiation of Echinodermata’, in Lipps, J. H. and Signor, P. W. (eds) *Origin and Early Evolution of the Metazoa*. Boston, MA: Springer US, pp. 375–398. doi: 10.1007/978-1-4899-2427-8_11.

Starr, T. N. and Thornton, J. W. (2016) ‘Epistasis in protein evolution’, 25, pp.

1204–1218. doi: 10.1002/pro.2897.

Stockhaus, J. *et al.* (1992) ‘Serine-to-alanine substitutions at the amino-terminal region of phytochrome A result in an increase in biological activity’, *Genes and Development*. doi: 10.1101/gad.6.12a.2364.

Su, Y. H. *et al.* (2009) ‘A perturbation model of the gene regulatory network for oral and aboral ectoderm specification in the sea urchin embryo’, *Developmental Biology*. Elsevier Inc., 329(2), pp. 410–421. doi: 10.1016/j.ydbio.2009.02.029.

Suga, H. *et al.* (2013) ‘The Capsaspora genome reveals a complex unicellular prehistory of animals’, *Nature Communications*, 4, pp. 1–9. doi: 10.1038/ncomms3325.

Summerton, J. (1999) ‘Morpholino antisense oligomers: The case for an RNase H-independent structural type’, *Biochimica et Biophysica Acta - Gene Structure and Expression*. doi: 10.1016/S0167-4781(99)00150-5.

Sweet, H. C., Gehring, M. and Ettensohn, C. A. (2002) ‘LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties’, *Development*.

Tavassoly, I., Goldfarb, J. and Iyengar, R. (2018) ‘Systems biology primer: the basic methods and approaches.’, *Essays in biochemistry*, 62(4), pp. 487–500. doi: 10.1042/EBC20180003.

Taylor, W. R. (1986) ‘The classification of amino acid conservation’, *Journal of Theoretical Biology*. doi: 10.1016/S0022-5193(86)80075-3.

Technau, U. (2001) ‘Brachyury, the blastopore and the evolution of the mesoderm’, *BioEssays*, 23(9), pp. 788–794. doi: 10.1002/bies.1114.

Technau, U. and Scholz, C. B. (2003) ‘Origin and evolution of endoderm and mesoderm’, 539, pp. 531–539.

Terazawa, K. and Satoh, N. (1995) ‘Spatial expression of the amphioxus homologue of Brachyury (T) gene during early embryogenesis of Branchiostoma belcheri’, *Development, Growth & Differentiation*, 37(4), pp. 395–401. doi: 10.1046/j.1440-169X.1995.t01-3-00006.x.

Thomas, M. C. and Chiang, C. M. (2006) ‘The general transcription machinery and general cofactors.’, *Critical reviews in biochemistry and molecular biology*. doi: 10.1080/10409230600648736.

Timme-Laragy, A. R., Karchner, S. I. and Hahn, M. E. (2012) ‘Gene Knockdown by Morpholino-Modified Oligonucleotides in the Zebrafish (Danio rerio) Model: Applications for Developmental Toxicology’, in, pp. 51–71. doi: 10.1007/978-1-61779-867-2_5.

Tippens, N. D., Vihervaara, A. and Lis, J. T. (2018) ‘Enhancer transcription: What, where, when, and why?’, *Genes and Development*. doi: 10.1101/gad.311605.118.

Tootle, T. L. and Rebay, I. (2005) ‘Post-translational modifications influence transcription factor activity: A view from the ETS superfamily’, *BioEssays*, 27(3), pp. 285–298. doi: 10.1002/bies.20198.

Tu, Q. *et al.* (2006) ‘Sea urchin Forkhead gene family: Phylogeny and embryonic expression’, *Developmental Biology*, 300(1), pp. 49–62. doi: 10.1016/j.ydbio.2006.09.031.

Tu, Q., Cameron, R. A. and Davidson, E. H. (2014) ‘Quantitative developmental transcriptomes of the sea urchin Strongylocentrotus purpuratus’, *Developmental Biology*, 385(2), pp. 160–167. doi: 10.1016/j.ydbio.2013.11.019.

Visel, A. *et al.* (2009) ‘ChIP-seq accurately predicts tissue-specific activity of enhancers’, *Nature*. doi: 10.1038/nature07730.

Walton, K. D. *et al.* (2006) ‘Genomics and expression profiles of the Hedgehog and Notch signaling pathways in sea urchin development’, *Developmental Biology*, 300(1), pp. 153–164. doi: 10.1016/j.ydbio.2006.08.064.

Walton, K. D. *et al.* (2009) ‘Hedgehog signaling patterns mesoderm in the sea urchin’, *Developmental Biology*. doi: 10.1016/j.ydbio.2009.04.018.

Wang, C. *et al.* (2001) ‘TAK1 is a ubiquitin-dependent kinase of MKK and IKK’, *Nature*. doi: 10.1038/35085597.

Waterhouse, A. M. *et al.* (2009) ‘Jalview Version 2-A multiple sequence alignment editor and analysis workbench’, *Bioinformatics*. doi: 10.1093/bioinformatics/btp033.

Wei, Z. *et al.* (2009) 'The sea urchin animal pole domain is a Six3-dependent neurogenic patterning center', *Development*. doi: 10.1242/dev.032300.

Wilkinson, D. G., Bhatt, S. and Herrmann, B. G. (1990) 'Expression pattern of the mouse T gene and its role in mesoderm formation', *Nature*, 343(6259), pp. 657–659. doi: 10.1038/343657a0.

Wilkinson, D. G., Bhatt, S. and Herrmann, B. G. (1997) 'Expression of T Protein in the Primitive Streak Is Necessary and Sufficient for Posterior Mesoderm Movement and Somite Differentiation', 58, pp. 45–58.

Wilson, V. and Conlon, F. L. (2002) 'The T -box family', pp. 1–7.

Wood, N. J. *et al.* (2018) 'Neuropeptidergic systems in pluteus larvae of the sea urchin strongylocentrotus purpuratus: Neurochemical complexity in a "Simple" nervous system', *Frontiers in Endocrinology*. doi: 10.3389/fendo.2018.00628.

Woollard, A. and Hodgkin, J. (2000) 'The Caenorhabditis elegans fate-determining gene mab-9 encodes a T-box protein required to pattern the posterior hindgut', *Genes and Development*, 14(5), pp. 596–603. doi: 10.1101/gad.14.5.596.

Xue, Y. *et al.* (2006) 'PPSP: Prediction of PK-specific phosphorylation site with Bayesian decision theory', *BMC Bioinformatics*. doi: 10.1186/1471-2105-7-163.

Yamada, A. *et al.* (2010) 'Highly conserved functions of the Brachyury gene on morphogenetic movements : Insight from the early-diverging phylum Ctenophora', *Developmental Biology*. Elsevier Inc., 339(1), pp. 212–222. doi: 10.1016/j.ydbio.2009.12.019.

Yamada, L. *et al.* (2003) 'Morpholino-based gene knockdown screen of novel genes with developmental function in Ciona intestinalis', *Development*, 130(26), pp. 6485–6495. doi: 10.1242/dev.00847.

Yasuoka, Y., Shinzato, C. and Satoh, N. (2016) 'The Mesoderm-Forming Gene brachyury Regulates Ectoderm-Endoderm Demarcation in the Coral Acropora digitifera', *Current Biology*. Elsevier Ltd., 26(21), pp. 2885–2892. doi: 10.1016/j.cub.2016.08.011.

Yuh, C. H., Dorman, E. R. and Davidson, E. H. (2005) 'Brn1/2/4, the predicted

midgut regulator of the endo16 gene of the sea urchin embryo', *Developmental Biology*. doi: 10.1016/j.ydbio.2005.02.034.

Zazueta-Novoa, V. and Wessel, G. M. (2014) 'Protein degradation machinery is present broadly during early development in the sea urchin', *Gene Expression Patterns*. doi: 10.1016/j.gep.2014.06.002.

Zhang, Y. *et al.* (2008) 'Model-based analysis of ChIP-seq (MACS).', *Genome biology*. doi: 10.1186/gb-2008-9-9-r137.

Zhao, Q. *et al.* (2014) 'GPS-SUMO: A tool for the prediction of sumoylation sites and SUMO-interaction motifs', *Nucleic Acids Research*. doi: 10.1093/nar/gku383.

Zhu, J. *et al.* (2018) 'Protein threading using residue co-variation and deep learning', in *Bioinformatics*. doi: 10.1093/bioinformatics/bty278.